Analysis of alternative Covid-19 mitigation measures in school classrooms: an agent-based model of SARS-CoV-2 transmission

M. J. Woodhouse[1], W. P. Aspinall[1,2], R. S. J. Sparks[1] & CoMMinS Project "COVID-19 Mapping and Mitigation in Schools"[3]

[1] *School of Earth Sciences, University of Bristol, Wills Memorial Building, Queens Road, Bristol BS8 1RJ, UK*
[2] *Aspinall and Associates, Tisbury SP3 6HF, UK*
[3] *Bristol Medical School (PHS), University of Bristol, BS8 2PS UK. https://commins.org.uk/*

Keywords: SARS-Cov-2; agent-based epidemiological model; Covid-19 in schools; random networks; structured expert judgement

1. Abstract

The SARS-CoV-2 epidemic has had major impacts on children's education, with schools required to implement infection control measures that have led to long periods of absence and classroom closures. We develop an agent-based epidemiological model of SARS-CoV-2 transmission that is applied to model infection within school classrooms, with a contact model constructed using random networks informed by structured expert judgement. Mitigation strategies to control infection are modelled to allow analysis of their effectiveness in supressing infection outbreaks and in limiting pupil absence. The model is applied to re-examine Covid-19 in schools in the UK in autumn 2020, and to forecast infection levels in autumn 2021 when the more infectious Delta-variant is dominant and school transmission is likely to play a major role in a new wave of the epidemic. Our results indicate that testing-based surveillance of infections in the classroom population with isolation of positive cases is a more effective mitigation measure than bubble quarantine both for reducing transmission in schools and for avoiding pupil absence, even accounting for insensitivity of self-administered tests. Bubble quarantine results in large numbers of pupils absent from school, with only modest impact of classroom infection. However, maintaining a reduced contact rate within the classroom has a major beneficial impact for managing Covid-19 in school settings.

2. Introduction

As an increasing proportion of people become vaccinated the spread of SARS-CoV-2 becomes concentrated mainly with unvaccinated persons. Children become of particular importance in these circumstances since the benefits, to children themselves, of vaccination is less clear-cut and controversial. While, at the time of writing, the UK government is yet to make a final decision on secondary age children, it seems less likely that primary age children will be mass vaccinated. Factors that influence the discussions include the observation that severe COVID-19 disease is very rare in children, while there are very small risks of adverse reactions to the vaccine. Thus, the efficacy of vaccination from the point of view protection for serious disease to an individual is equivocal. On the other hand, schools with largely unvaccinated populations may act as environments for spreading SARS-CoV-2, with the potential for development of new variants. From a public health point of view then schools are, potentially, a significant reservoir of infection. Measures to mitigate against transmission can also be very disruptive to learning and to the economy if large numbers of pupils and their families are required to self-isolate or have their work compromised by childcare responsibilities. In the latter context there is an urgent need to understand what policies should be pursued that reduce transmission but at the same time minimise disruption of education and collateral effects.

In this study we have developed a basic stochastic model for estimating the likelihood of the transmission of SARS-CoV-2 occurring in primary schools. The model is based on a discrete agent-based compartmental epidemiological model and involves an empirical representation of the probability of transmission due to close contacts within class-sized interactive networks. With our approach, sets of several classrooms collectively can be assessed as a way of quantifying transmission probabilities in individual schools. Statistical representation of close contacts, defined in [1] to be a face-to-face contact within 1 metre for at least 5 minutes, between children and adults within school settings is derived from a study of 36 primary schools in England using teachers' expert judgements, formally elicited in Spring/Summer 2020 [1]. That study included a comparison of anticipated contact rates in schools taking mitigation measures in the following September 2020, such as formation of bubbles, reduced class sizes and other social distancing rules, with previous, normal contact rates in 'pre-Covid' times.

The model developed here is compared with data on school attendances, self-isolation of children who are sent home due to Covid outbreaks and prevalence of Covid in both schools and communities. These comparisons enable us to check that the model produces plausible results. However, the main purpose of our modelling approach is to compare the effects of different mitigation strategies on infection transmission rates within schools, including sending '*bubbles*' or whole classes home for self-isolation periods, sending home only those pupils or adults thought to have become infected, or using testing to identify infections.

At the centre of our approach is a time-stepping agent-based epidemiological model of SARS-CoV-2 infection transmission in a classroom. The model we have developed combines a discrete compartmental epidemiological framework with a random network for daily person-to-person contacts within the classroom. In the basis case, the classroom population consists of a single teacher, a set number of pupils, and a small number of classroom teaching assistants. In relation to infection transmissions, the modelled classroom population is assumed to be isolated from other classes and persons within the school, so the present model does not include interactions with other people in the school. This is a major assumption and is best suited to primary school settings where classroom groups can be most effectively separated from one another as a part of infection risk management; this was a mitigation instituted in primary schools in England that were open for vulnerable children and those of critical workers in March 2020 and as schools in England reopened to selected age groups in June 2020. The simplifying assumption is also justified by evidence for limited mixing across different classes and year groups in UK primary schools [2].

As the model advances through time, it includes a daily chance of seeding infection within one or more of the classroom occupants from interactions with the outside community; this infection probability is a function of community incidence rate. We describe this aspect of our model in more detail below.

We then apply our model to the Autumn term (Term 1) in 2020 when UK schools first reopened to full classes. Using estimates of SARS-CoV-2 infection prevalence and the incidence rate during Term 1 2020, we illustrate the application of our model and show that it can simulate numbers of infections that are in good agreement with available data. We then apply the model over this period with different mitigation measures applied and analyse the effect of these strategies on the transmission of SARS-CoV-2 within schools. We then apply our model to forecast infection in schools for the upcoming Term 1 in 2021, where schools may play an important role in a third-wave of the SARS-CoV-2 epidemic, particularly in the face of the Delta variant and exclusion of children from vaccination programmes.

3. Methods

Epidemiological model

The progression of infection is described using a variation on the Susceptible—Exposed—Infectious—Quarantined—Recovered (*SEIQR*) compartmental epidemiological model, which includes explicit time variation in the transmissibility of SARS-CoV-2. Using evidence derived from large datasets on the timing of secondary infections with the onset of symptoms of primary cases [3], we model infectivity through time-dependent transmission probability of an index case with a contact. The transmission probability increases during the incubation period from zero at the time of infection, peaks slightly before the time of symptom onset (for symptomatic cases), and subsequently decreases over time [3]. It is inferred from data [3] that there is a distribution of the duration of incubation period of SARS-CoV-2 and that the duration of pre-symptomatic infectious period is related to the duration of the incubation period, with individuals who have long incubation periods tending to have an earlier and longer lasting pre-symptomatic infectious period. Variants of the virus could alter these distributions.

These time-dependencies are encoded in an expression for the relative probability of transmission in a contact within the class that is a function of time (t, in days) since symptom onset [3], denoted here by $p_r(t|t_i)$, which depends on the incubation period of the infected individual (t_i, in days). For the distribution of the incubation period we use a log-Normal distribution, $t_i \sim \log N(1.63, 0.5)$ [4]. For the relative probability of transmission, we adopt the parametric model determined from data by [3],

$$p_r(t|t_i) = \alpha(1+1/\alpha)^{\alpha+1} \times \begin{cases} \dfrac{e^{-(\tau t/t_i-\mu)/\sigma}}{(1+e^{-(\tau t/t_i-\mu)/\sigma})^{\alpha+1}}, & -t_i < t < 0, \\ \dfrac{e^{-(t-\mu)/\sigma}}{(1+e^{-(t-\mu)/\sigma})^{\alpha+1}}, & t \geq 0, \end{cases}$$

where the parameters $\mu = -4$ days, $\sigma = 1.85$ days, $\alpha = 5.85$ are fitting parameters, and $\tau = 5.42$ days is the mean incubation period [4]. Note, this function is not a probability density function, rather it is applied as a time-varying weighting to the transmission probability in a contact, and the pre-factor is specified so that $\max_t p_r(t|t_i) = 1$.

The time variation of the relative transmission probability is illustrated in Figure 1 for an unusually short ($t_i = 2,\ P(t_i < 2) < 0.05$), a typical ($t_i = 5,\ P(t_i < 5) = 0.48$) and an unusually long (t_i=12, $P(t_i > 12) < 0.05$) incubation period.

The absolute probability of transmission in a contact is found by scaling $p_r(t|t_i)$ by a fixed peak transmission probability, which differs for adults and children [5-7], for symptomatic or asymptomatic cases [8], reflects mitigation measures, and is altered to model differences in transmissibility for SARS-CoV-2 variants.

The classroom population can comprise both symptomatic and asymptomatic individuals, distinguished as these who will (respectively, will not) develop identifiable symptoms if they are infected with SARS-CoV-2. The specification of symptomatic and asymptomatic individuals occurs on initialization of the model by random Bernoulli draws with a prescribed probability. For asymptomatic cases, we adopt the same time-varying relative probability of transmission by specifying a notional incubation time (drawn from the log-Normal distribution), but the absolute probability of transmission in a contact is reduced relative to symptomatic cases.

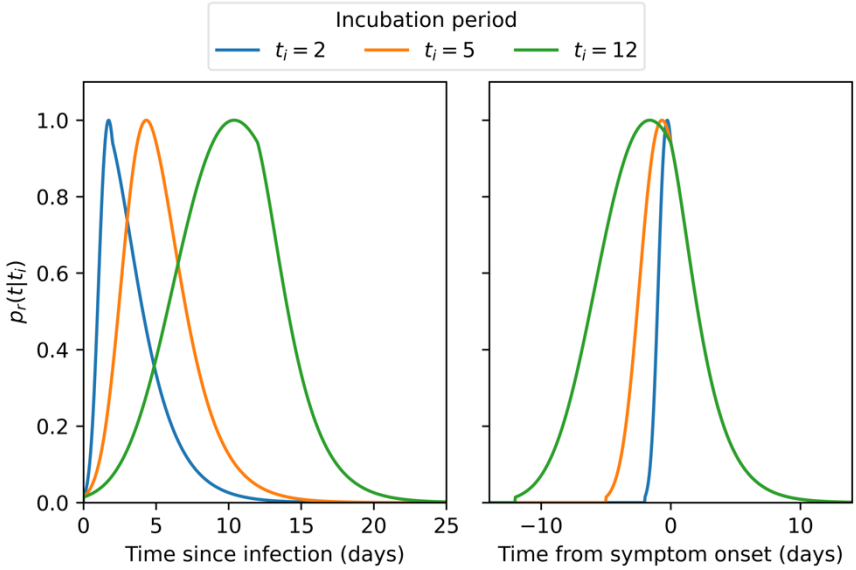

Figure 1. Time variation of the relative probability of transmission from an infected individual with incubation period t_i. Probability of transmission increases from the time of infection, peaks close to the time of onset of symptoms, and then decreases over the duration of infectivity. Individuals with long incubation periods have longer pre-symptomatic infection periods, but following symptom onset the decrease in transmission probability is independent of t_i.

Symptomatic individuals become unwell following the incubation period. The duration of symptoms varies for adults and children [e.g., 9, 10] with adults typically having longer lasting symptoms (median duration of 11 days) compared to children (median duration of 5 days for children aged 5—11 years [10]). However, there is pronounced variation around the median duration which we model using a log-Normal distribution. This is a two-parameter distribution, so it is specified by using the reported proportion of people having symptoms extending beyond 28 days in [9, 10] (13.3% for adults and 3.1% of primary school aged children).

To accommodate transmission from pre-symptomatic and asymptomatic individuals, we use a compartmental model with the following compartments: Susceptible (*S*), Exposed (*E*), Unwell (*U*), Quarantined (*Q*), and Recovered (*R*). The Susceptible population are those who have not become infected with the SARS-CoV-2 virus. On becoming infected, an individual moves to the Exposed compartment and begins a period of incubation during which there is no external indication of infection, but where there is a possibility of infection transmission in contacts. For symptomatic individuals the incubation period ends with the individual experiencing symptoms and moving into the Unwell compartment. At this stage we assume that symptoms of infection would be recognized by the infected individual or observed by others and, if such symptoms are detected or confirmed, usually by testing, then an Unwell individual will be Quarantined. Individuals remain quarantined for a minimum of the quarantine period, and for as long as they are unwell. However, asymptomatic cases have no symptoms, or symptoms so mild as to be unobserved, so these persons do not migrate to the Unwell category. The Recovered compartment collects infected individuals following their period of being Unwell, or for the duration of the infectious period for asymptomatic cases, at the which stage they are assumed to have acquired immunity from further SARS-CoV-2 infection.

An additional ingredient of the population dynamics is that quarantined teachers must be replaced by a substitute teacher, drawn from the outside population. We adopt a discrete-time stochastic modelling framework, with updates to the compartmental model occurring each day.

Initialization of the model places the classroom members into the *SEUQR* compartments. A Bernoulli trial is performed for each individual using a specified community infection prevalence, with '*failure*' resulting in the individual placed into the Susceptible compartment. A '*success*' in the Bernoulli trial corresponds to an individual with infection on initialization, so random draw is made to determine their elapsed time of infection and they are placed into the appropriate E, U or Q compartments on this basis.

Classroom transmission network

In our model there are two pathways for susceptible members of the classroom to become exposed: contact with an infectious member of the classroom, or through interactions outside the classroom (referred to here as '*community transmission*'). To model the infection from an out-of-classroom interaction, we begin each daily time step with a Bernoulli trial for each susceptible classroom member. The probability of any susceptible classroom individual being exposed to community transmission is specified using an estimate of the daily incidence rate of new infections at the appropriate time. This approach overestimates community transmission as the individual is in school for much of the day, and classroom transmissions will contribute to the community incidence rate estimates. However, without detailed estimates of incidence rates in different settings, it is difficult to deconvolve these effects, and our approach is a pragmatic comprise, recognizing that time in school represents approximately 26% of waking time in any school week, and our model does not explicitly include effects that might promote incidence of infection connected to schools (such as increased social mixing at the start and end of the school day).

Transmission of infection within the classroom requires a more sophisticated model, and here we use a random network model of contacts within classroom members. The number of contacts for each individual is specified from stochastic draws from the distribution of daily classroom contacts [1]. Thus, each class member has their own number of daily contacts with others in the classroom.

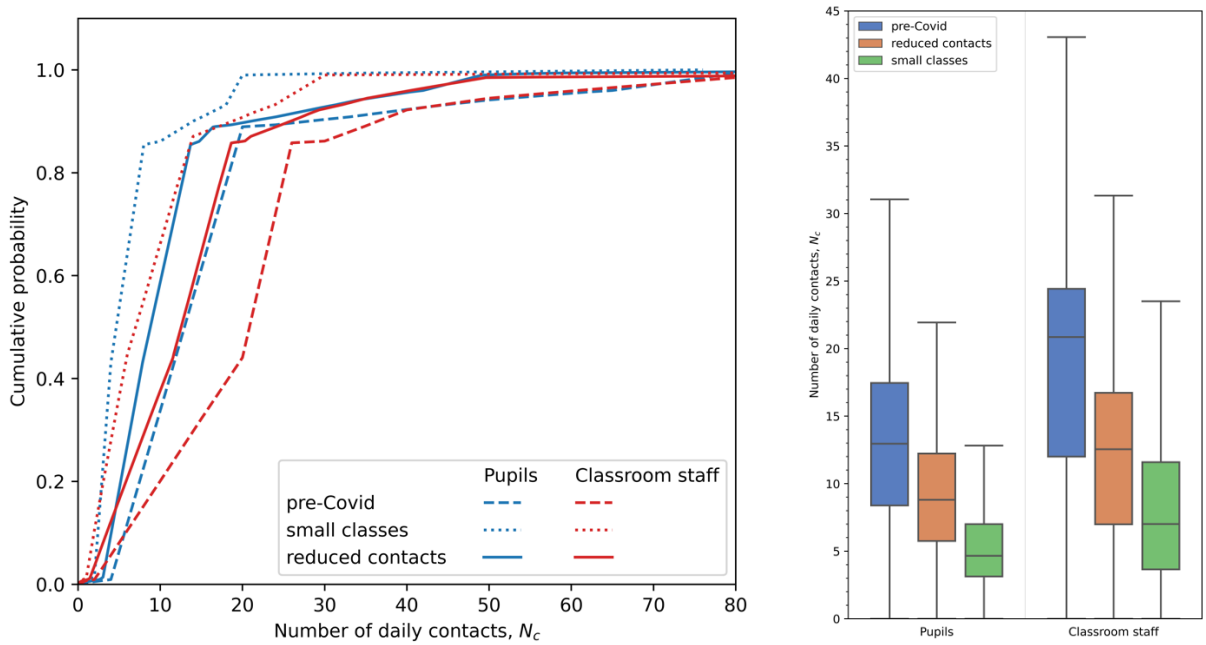

Figure 2. a) Empirical cumulative distribution functions (eCDFs) for the number of daily contacts in primary school classrooms. The eCDF derived from Structured Expert Judgement for pupils (blue) and classroom staff (teachers and teaching assistants; red) are shown for pre-Covid times (dashed lines) and following opening of schools to small numbers of children in June—July 2020 (dotted lines). The cumulative distribution function constructed from a weighted combination of the pre-Covid and small classes eCDFs are shown (solid lines) b) Box and whisker plots (with outliers removed) comparing the number of daily contacts for pupils and classroom staff in pre-Covid times, for small classes, and modelled for full classes under conditions to reduce contacts.

It is known that contact patterns in classrooms can be highly variable, affected by both individual and classroom behaviours [1]. Additionally, behavioural mitigation measures were instituted in March 2020 to reduce contacts within schools. To model the number of contacts, we use the results of a Structured Expert Judgement (SEJ) elicitation study [1] conducted in spring/summer 2020 in the UK. SEJ uses the collective knowledge of experts (here school teachers) to estimate quantities of interest and their uncertainties, and can be performed rapidly during crises to provide essential information to models and policy makers [11]. Experts' responses to a set of test questions with knowable results are used to score experts in terms of their individual statistical performance and their target item judgements are then combined jointly, using these performance scores, to derive a pooled 'Decision

Maker' [11]; that the latter represents associated group-wise judgement uncertainties through empirical cumulative distribution functions (eCDFs) [12].

In our model, empirical cumulative distribution functions for the number of contacts in classrooms are those derived from pooled expert responses in an elicitation for primary school pupils and classroom staff [1]. That study determined different contact rates in: (a) the pre-Covid classroom, and then (b) following implementation of bubbles and other measures to reduce the number of contacts [1]. The corresponding eCDFs are illustrated in Figure 2a, showing substantial reduction in the number of contacts once mitigation measures are imposed.

A problem here is that the contact data estimated by SEJ in [1] were obtained under the unique circumstances of June and July 2020 when between 30% and 40% of children had returned to school and the class sizes were greatly reduced, making social distancing measures much easier to implement. However, with full return to school we would not expect the same reduction in contact rates. In study [1] teachers were asked to estimate whether the contact numbers would be affected. Their collective view (Table 7 in reference [1]) is that the contact numbers would be between their estimates in normal pre-Covid times and in new normal times.

To model the reduction in the number of contacts in fully-opened schools, we construct a cumulative distribution function (denoted by $P_r(N = n)$) as a weighted linear combination of the eCDF in pre-Covid classrooms in normal times (denoted by $P_n(N = n)$) and the eCDF in the small classrooms of June—July 2020 (denoted by $P_s(N = n)$), with $P_r(n) = wP_n(n) + (1 - w)P_s(n)$ where w is the weighting of the pre-Covid contact distribution. Contact rate distributions are constructed in this way for both pupils and teachers, with differing weights. In [1] teachers expected that daily pupil contacts would be half-way between the small classroom and normal pre-Covid values when classrooms become more highly occupied, and we therefore take a weight $w = 0.5$. However, teachers considered that their own numbers of daily contacts could be more strongly reduced in the fully open schools, so we take $w = 0.4$ to place greater weight on the small classroom distribution.

At the start of each school day in the modelled sequence, we build a stochastic network of contacts for each class member who is not quarantined. The network is constructed using the '*Configuration Model*' approach [e.g. 13], allowing multiple edges but removing self-loops through a sequence of random '*rewirings'* (i.e. self-loop edges are randomly switched until a valid configuration is achieved). Preserving multiple edges allows us to model repeated contacts between individuals. The Configuration Model produces a random graph with a specified degree distribution in which nodes of the network are linked randomly using specified numbers of edges. This means that gregarious individuals with high numbers of contacts are more likely to be connected to other highly contacting individuals. However, the contact network is rebuilt each day, to reflect changing contact patterns and the removal of class members into quarantine. The Configuration Model requires an even number of connections; in cases where an odd number of edges is generated by the stochastic initialization of the degree distribution or from removal of nodes to quarantine, then the degree of a randomly selected node is increased by one. On non-school days, there are no classroom contacts, but community transmission can occur, and the agents continue to progress through the *SEUQR* compartments.

If contact occurs between an infectious individual with another who is susceptible, there is the possibility of transmission. This is modelled as a random Bernoulli trial, with differing probabilities of transmission depending on whether the infectious individual is symptomatic (higher transmission probability) or asymptomatic (lower transmission probability), and with transmission probability varying in time. Note, this approach combines the infectiousness of the index case with the susceptibility of their contact into a single probability of transmission. This is a simplifying assumption in the model, but removes individual susceptibility to SARS-CoV-2 for the population in the model.

An additional requirement is that the classroom must have a teacher present. If the teacher is quarantined due to illness or a positive test result, a temporary replacement is added to the classroom

population. The substitute teacher has their own number of daily contacts, drawn from the SEJ distribution, and their stage in the *SEUQR* progression is determined by Bernoulli trial with the infection prevalence at the time of replacement, noting that a valid substitute teacher cannot be showing symptoms (i.e., Unwell) nor in Quarantine, but could bring infection into the classroom population. If the permanent teacher is released from Quarantine, or their time of illness elapses (i.e., they progress to Recovered), then they are returned to the classroom and the substitute teacher is removed. Should the temporary teacher become Unwell or Quarantined, they are replaced by a different, substitute teacher.

Examples of three random contact networks produced using the Configuration Model are illustrated in Figure 3 for a small classroom of ten pupils (enumerated from 1 to 10) and a single teacher (labelled as 'T'), simulated over three days. The degree distribution is fixed; pupils 1 and 2 are the most gregarious with five contacts each day, followed by pupil 5 with five contacts on days 1 and 3 but four contacts on day 2. There are duplicated contacts (i.e., multiple edges) between pupils in each network, which present two opportunities for transmission between persons if one of the individuals is infectious and the other susceptible (which occurs in Figure 3c between pupils 1 and 2).

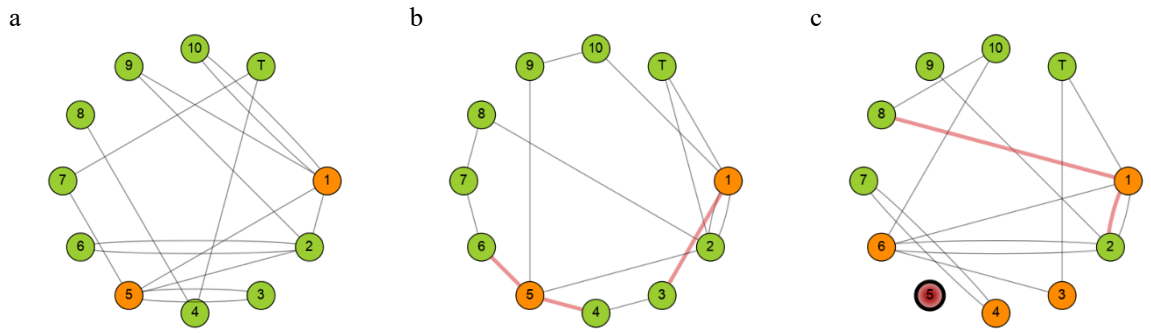

Figure 3. An example of infection transmission within a random updating contact network in a small classroom of ten pupils and one teacher. The degree distribution is fixed, and the edges are determined using the Configuration Model with self-loops removed by random rewiring. Multiple edges are allowed, as illustrated in a *where e.g. there are two edges between nodes 1 and 10. The panels represent a progression over three days. Green nodes indicate Susceptible class members, orange nodes indicate Exposed members, and red nodes indicate Unwell individuals. Nodes that are boldly edged (node 5 in* c*) are quarantined. Red coloured edges indicate infection transmission between infectious and susceptible members. Note the parameters for infection transmission in this example are not representative of SARS-CoV-2.*

Figure 3 also illustrates the transmission of infection across the network, with a daily progression from panels a to b to c. Note that the parameters related to infection transmission are here taken to be much larger than reasonable to model SARS-CoV-2 and have been chosen simply to illustrate the way the model updates. On the first day (Figure 3a), there are two exposed individuals (pupils 1 and 5), and their contacts represent potential pathways for infection, with actual transmission determined by Bernoulli trials. In this example, the random trials result in no transmissions. However, on the second day (Figure 3b) the two exposed pupils do transmit the infection across some of their contacts; of the five contacts with pupil 1, there is one transmission to pupil 3, whereas two contacts of pupil 5 result in transmission to pupils 4 and 6. Therefore, on the third day, pupils 3, 4 and 6 are progressed from the Susceptible into the Exposed compartments. Further infection transmission occurs on day 3, with two contacts of pupil 1, but, by chance and fortuitously, there is no infection transmission in contacts of the other exposed individuals. Additionally, on day 3, the pupil 5 has developed symptoms and is progressed into the Unwell compartment. In this example, the Unwell individual is Quarantined away from the classroom and has no connections with other members of the classroom.

Vaccination uptake and effectiveness

As vaccination programmes continue, there is a need to include the effects of vaccine on SARS-CoV-2 infection, transmission and symptoms. In the UK, vaccination began in October 2020 with high-risk groups and has proceeded downwards through age-groups. Therefore, in modelling school infection

in Term 1 2020, there are very low levels of vaccination for pupils and teachers. However, vaccine uptake has been high, so that by the end of the 2020—2021 school year 80% of working age people are estimated to have received two vaccine doses [14]. Primary aged children are unlikely to have received vaccine unless they have additional risk factors. To include vaccination in our model, on initialization of the class members, we perform Bernoulli trials using estimates of the vaccination uptake proportion.

We model three effects of the vaccine: (i) a reduction of the probability of a vaccinated individual becoming infected; (ii) a reduction of the probability of an infected vaccinated individual transmitting the infection in a contact; (iii) a reduction of the probability of an infected vaccinated individual developing symptoms. The different available vaccines have different effectiveness in each of these [15, 16]. Here we use fixed values for scaling factors applied for the reduction in probability of infection, reduction in probability of transmission, and reduction in probability of symptoms for vaccinated members of the population. Therefore, for susceptible vaccinated individuals, their probability of becoming infected by both community transmission and classroom contact is reduced by a '*Vaccine Susceptibility Factor*'. For an infected vaccinated individual, their probability of onward transmission in a contact is reduced by a '*Vaccine Transmission Factor*', and their probability of developing symptoms is reduced by a '*Vaccine Symptoms Factor*'. These are assigned plausible values based on [15-17] (Table 1).

Mitigations

To understand the effectiveness of mitigation measures in reducing SARS-CoV-2 transmission and infection prevalence within the classroom population, we consider several '*hard*' and '*soft*' approaches. Soft measures include enhanced cleaning and hygiene, as well as the reduction of mixing of separate classrooms (e.g., through staggered entry, breaks and lunch times; multiple and separated entry points to the school, etc.). We do not model these explicitly as the efficacy of some of these measures is limited [18] and is not well known and would substantially increase the complexity of the model and its interpretation. Instead, we apply soft mitigation factors that reduce the absolute probability of transmission in a contact by a fixed factor and specify different values for adults and children.

Hard mitigations correspond to isolation of classroom members away from the school and result either from the identification of symptoms or from testing surveillance to detect pre-symptomatic infection or their contacts. We have discussed above the removal of symptomatic individuals from the classroom as they reach the Unwell stage of the *SEUQR* progression. In our model, we also allow for two other hard mitigations: '*bubble quarantine*' of the entire class, and '*regular rapid testing*' of all individuals.

The '*bubble quarantine*' approach is severe, with the entire class placed into Quarantine for a specified time period if there is a confirmed infection case in the class (e.g., if a classroom member becomes unwell). During this time, there are no classroom contacts. However, classroom members may still acquire infection through community transmission and continue to progress through the S, E, U, R compartments. For Susceptible members in quarantine, the probability of becoming infected through community transmission is likely to be reduced as their contacts will be substantially reduced. However, the as infection transmission within households is the primary pathway for infection [19], the reduction in community transmission is likely to be small, and for simplicity we do not include in our model as it would introduce a further highly uncertain parameter. Any individual who becomes Unwell during Quarantine restarts their quarantine period. On completion of Quarantine all classroom members (except those who are Unwell) are returned to the classroom. Note, there may be asymptomatic or pre-symptomatic infected individuals returned to the classroom following bubble quarantine.

An alternative and less severe approach is '*regular rapid testing*'. This surveillance method adopts the regular lateral flow testing of all the classroom population on specified days. The rapid testing is capable of identifying asymptomatic and pre-symptomatic infection, allowing early quarantine by

advancing members from the Exposed compartment to Quarantine. However, all tests have the potential to return misleading results, which may be particularly acute when administered by untrained people [20]. Therefore, we model the application of rapid testing using a stochastic approach. On each test day, each classroom member undergoes a Bernoulli trial: for Exposed individuals who have the infection, the trial adopts the rapid test sensitivity (the true positive rate, i.e., the test probability of correctly detecting infection in an infected subject); for Susceptible individuals who are not infected, the trail adopts the rapid test false positive rate (i.e., the test probability of incorrectly detecting infection in an uninfected subject $= 1 -$ specificity). A '*success*' in each trial corresponds to a positive test result and leads to the individual moving into Quarantine. However, for uninfected individuals this is an incorrect test result, and the Susceptible member is unnecessarily removed from the classroom.

The sensitivity of rapid testing in the UK can be high if administered correctly [20]. However, in studies of lateral flow tests self-administered by untrained members of the public, the sensitivity can be as low as 0.58 [21]. We adopt this value here. We have not found similar values for the test specificity for self-administered lateral flow tests, so adopt the reported value of 0.9968 despite our expectation that this value is inflated.

We consider a further approach to reduce the number of classroom members unnecessarily quarantined by false positive rapid tests through the use of confirmatory PCR testing. The PCR tests are more accurate (high sensitivity and high specificity) laboratory testing, so are not well suited for regular testing of large populations. Here we allow for them to be used to confirm (or overturn) a positive rapid test result. The PCR test results do not appear immediately; rather, there is a lag while the test is processed, during which time the test-positive subject remains in Quarantine. The sensitivity and specificity of the PCR tests are reported to be much higher than rapid lateral flow tests [22].

Model parameters and stochastic ensemble

The parameters in the model are presented in Table 1. Many of them are uncertain; where possible we have adopted published values as noted in Table 1 (although many studies have not completed peer-review at the time of writing), and otherwise we have taken reasonable values. We have not attempted to 'fit' parameters to data, but such a study would be valuable. Therefore, our reported values are best used as comparison between the different scenarios we simulate rather than hind- or forecasts.

There are several stochastic components in the model, which we also summarise in Table 1, and therefore we must perform stochastic ensemble simulations of sufficient size to explore the space of possible trajectories of the infection dynamics within a classroom. In this study, we take ensembles with 100,000 realizations of isolated classrooms. Note, in school year 2020—2021 there are 156,843 state primary school classrooms in England [23]. Ensembles of 100,000 samples of our hypothetical class of 30 pupils provide indicative traits and trends in distributional variability; if we invoke the ergodicity assumption, while admitting national schools classes, *en masse*, will have a variety of pupil numbers and will have other factors in play (inner city -v- rural; economic differences, etc), then our 30-pupil class infection distributional profile can be adopted as first order representative of all real classes *faute de mieux.*

School collections

Although we model SARS-CoV-2 in a single isolated classroom, the level of infection in larger collections of classrooms (e.g., classes in a school, and schools together in a local authority area) are often the most pertinent issues for decision makers. We therefore model a school as a collection of classrooms, and several schools collected together, through sampling with replacement from a stochastic ensemble of simulations of independent individual classes. Specifically, we generate an ensemble of simulations and randomly select realizations from this ensemble to produce a collection of classrooms to represent a single school, and a collection of such schools to model e.g., a local authority area or academy network. An ensemble of such classroom collections is readily derived by

resampling with replacement from the single-classroom ensemble so that statistically meaningful metrics can be derived. We take this approach when comparing mitigation strategies.

Table 1. Stochastic and fixed parameters of the model.

Parameter	Distribution	Update
Initial infection in each class member	Bernoulli with community prevalence of infection	Set on initialization for class members using estimates of community prevalence. Updates on change of teacher using community prevalence time series.
Probability of infection from community	Bernoulli with community incidence rate	Daily, using time series of community incidence rate from observations or large-scale forecast models
Number of daily contacts for each class member	Non-parametric CDFs derived from SEJ for pupils and classroom staff. CDFs for pre-Covid and 'new normal' reduced contacts	Set on initialization for pupils. Updates on change of teacher. Contact network altered daily.
Symptomatic status	Bernoulli with specified probabilities symptomatic adults and children. Default value of 0.8 for both.	Set on initialization. Updates on change of teacher.
Duration of incubation period, t_i (note, asymptomatic cases have a notional t_i)	Drawn from log-Normal distribution [4]; $t_i \sim \log N(1.63, 0.5)$	Set on initialization.
Duration of unwell period, t_w (note, asymptomatic cases have a notional t_w)	Drawn from log-Normal distributions for children and adults [9, 10]; adults: $t_w \sim \log N(2.40, 0.840)$ children: $t_w \sim \log N(1.61, 0.923)$	Set on initialization. Updates on change of teacher.
Probability of contact transmission (base values for original wildtype SARS-CoV-2)	Adults / Children Symptomatic: 0.035 / 0.0175 Asymptomatic: 0.0245 / 0.01225	Fixed values. 50% reduction for children relative to adults [5, 6]; 30% reduction for asymptomatic relative to symptomatic infection.
Soft mitigation factor	Pupils: 0.75 Adults: 0.6	Fixed values
Rapid testing	Values from [21]. Sensitivity: 0.58 Specificity: 0.9968	Fixed values
PCR testing	Values from [22]. Sensitivity: 0.90 Specificity: 0.999	Fixed values
Weighting of pre-Covid contact rates to small classes contact rates, w	Pupils / Classroom staff Pre-covid: 1 / 1 Reduced contacts: 0.5 / 0.4 Small classes: 0 / 0	Fixed values
Probability of vaccinated adult	Term 1 2020: 0 Term 1 2021: 95% [14]	Fixed values
Probability of vaccinated child	Term 1 2020: 0 Term 1 2021: 1% [14]	Fixed values
Vaccine transmission factor	0.60 [15-17]	Fixed values
Vaccine susceptibility factor	0.40 [15-17]	Fixed values
Vaccine symptomatic factor	0.20 [15-17]	Fixed values

4. Retrospective analysis and model validation

The beginning of the 2020—2021 school year in September was the first time that there was 'full' attendance in schools since the start of the SARS-CoV-2 epidemic in the UK. The prevalence of infection in the UK in September 2020 was relatively low after the aggressive mitigation measures put in place in spring and early summer, but relaxation of restrictions over the traditional school summer holiday period in August resulted in increasing incidence of infection across the UK as schools reopened in September, a trend that persisted through Term 1 (Tuesday 1 September to Friday 23 October 2020) as shown in Figure 4. Data from the Office for National Statistics [24] gives infection prevalence of 0.09% for adults and 0.11% for children (age 2 to school year 6) on 1 September 2020, increasing by about an order of magnitude, to 1.10% for adults and 0.87% for children, by 23 October 2020. The incidence rate (per 10,000 population per day) was 0.74 on 1 September, peaked at 8.44 on 21 October 2020, falling slightly to 8.41 on 23 October 2020. At this time, vaccine uptake was low (the national vaccination programme had not started in September and was targeted at high-risk groups through October 2020) so we do not include vaccination in our simulations.

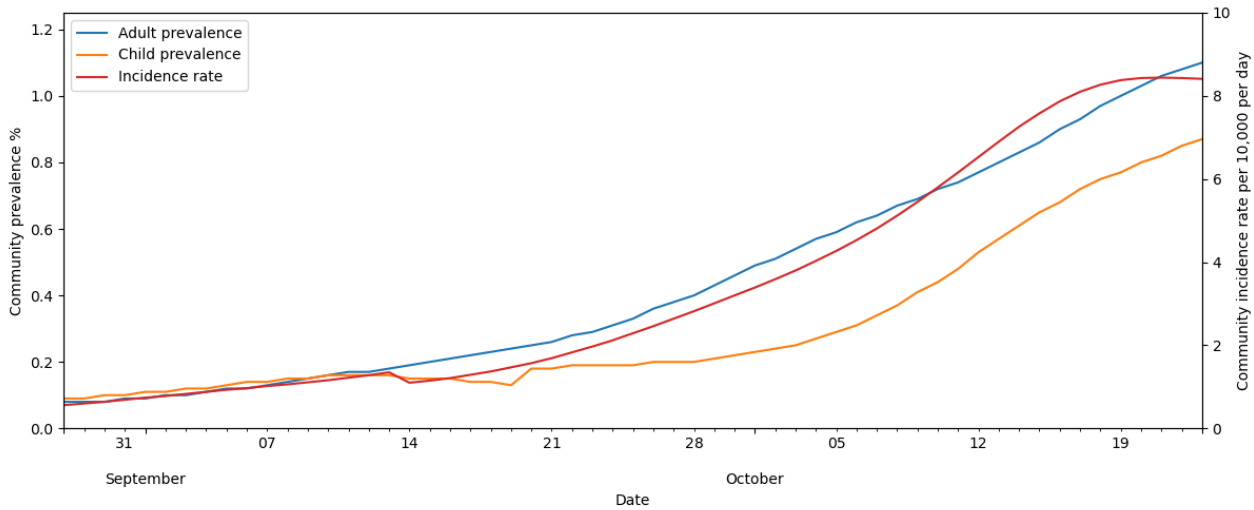

Figure 4. Time series of Covid-19 infection prevalence for adults (blue) and children (orange) and the community incidence rate (red) over Term 1 (Tuesday 1 September to Friday 23 October 2020). Data from Office for National Statistics [24].

We model a 'typical' primary school classroom, consisting of 30 pupils and a teacher, over Term 1. An ensemble with 100,000 realizations is computed; we consider this number of samples, for our hypothetical class of 30 pupils, can be regarded as providing indicative traits and trends in infection distributional variability in most primary school classes. While acknowledging schools' classes nationally will have a variety of pupil numbers and will have other factors in play (inner city -v- rural; economic differences, etc), by invoking ergodicity, we adopt our 30-pupil class infection distributional profile as being first order representative of all real classes nationally, *faute de mieux*. On this basis, we expect that findings from the realizations will scale to transmission at a national level. From national statistics [23] (see also [25]), there were 156,843 classes (4.18 million pupils) in English state primary schools in 2020. At this time there were no 'hard' mitigation measures, and pupils were only asked to self-isolate if they (or a close contact) developed symptoms. However, soft mitigations including enhanced cleaning and reduction of classroom contacts were in place. We therefore do not incorporate hard mitigation measures in the model simulations but build our contact networks using the reduced contacts eCDFs from SEJ, while presuming pupils are quarantined if they become unwell.

In Figure 5 we illustrate the dynamics of classroom transmission over Term 1 by plotting the timeseries of the number of infected pupils in a classroom for each of the ensemble members. The ensemble members have been ordered by the maximum number of infected pupils, which peaks at 18 across all ensemble members. A slight majority of realizations (53%) have no infected pupils across the whole of Term 1. The numbers of infected pupils increase towards the end of Term 1, following the trend in

the community incidence rate and, we infer, reflects seeding of infection in the classroom largely from outside community transmission. Figure 6 shows the number of infections within the ensemble that results from either community (Figure 6a) or classroom transmission (Figure 6b), and their correlation with the total number of infections.

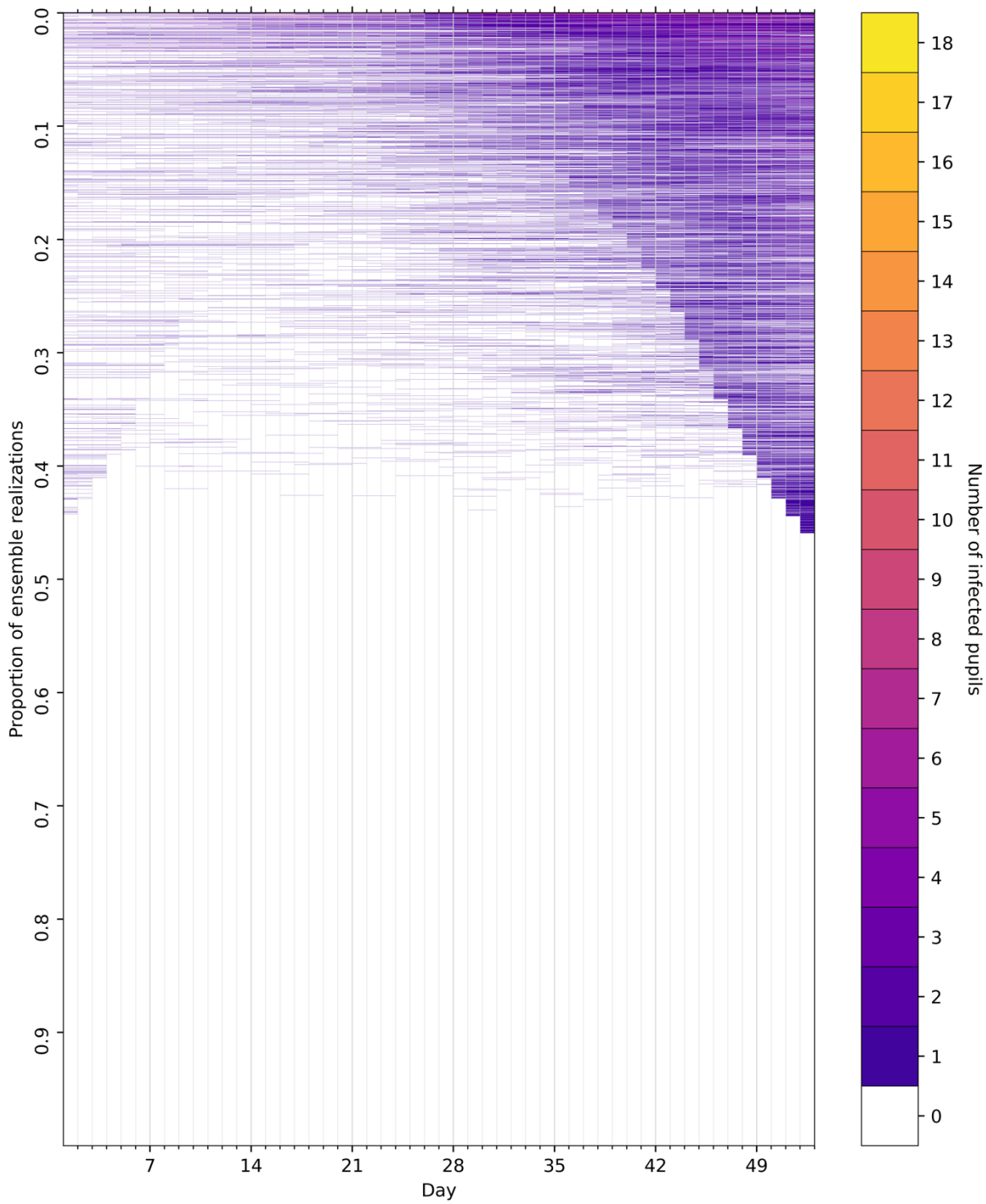

Figure 5. Time series of the number of infected pupils for each realization in an ensemble of 100,000 stochastic simulations of an isolated classroom in Term 1 of 2020. Each row of the heatmap corresponds to an ensemble member, with colours denoting the number of pupils in the classroom with Covid-19 infection on each day. The ensemble realizations have been sorted by the sum of the number of infected pupils over all days. Colours denote the number of infected pupils in the classroom for each day of the simulation.

When the total number of infections is up to six, we find that most infection occurs due to community transmission (Figure 6a), but beyond six infections this correlation is broken and in all these simulations (with infections >5) the number of classroom infections significantly exceeds the number

of infections expected in the community. These cases could pragmatically be defined as 'clusters' or 'outbreaks' due to in-class transmission being dominant. Current Department for Education (DfE) contingency operational guidance for schools [26] defines two thresholds for "extra action" under specific setting conditions and in relation to 'close mixing' groups within schools: "5 children, pupils, students or staff, who are likely to have mixed closely, test positive for COVID-19 within a 10-day period; or • 10% of children, pupils, students or staff who are likely to have mixed closely test positive for COVID-19 within a 10-day period". These criteria relate to two completed rounds of tests in an 'asymptomatic test sites' regime, initiated only following full return to school; it is not clear if there are criteria for similar actions due to in-school outbreaks under other circumstances.

In our ensemble there are relatively few realizations (1.1%) with total infections exceeding six, suggesting such outbreaks are relatively uncommon. Larger outbreaks of ten infected pupils, do not occur in our ensemble unless classroom transmission occurs. We find outbreaks with ten or more infected pupils occur in 0.1% of the simulations, while very large outbreaks of 15 or more pupils (i.e., at least half of our simulated class are infected) occur in only 14 (0.014%) of the realizations.

Open data on which to compare our results is scarce, but some summary statistics on school attendance have been published through the pandemic by the UK Department for Education (DfE) [27]. These statistics report that on 22 October 2020, about 20% of state-funded primary schools reported one or more pupils who were self-isolating. While it is difficult to compare this summary statistic with our stochastic simulations, as we do not know how many classes had pupils in self-isolation, the value compares favourably with our ensemble which has 18,778 (19%) of the realizations having one or more pupils quarantined on this day of the simulation. Based on the simulations we also expect about 2470 outbreaks.

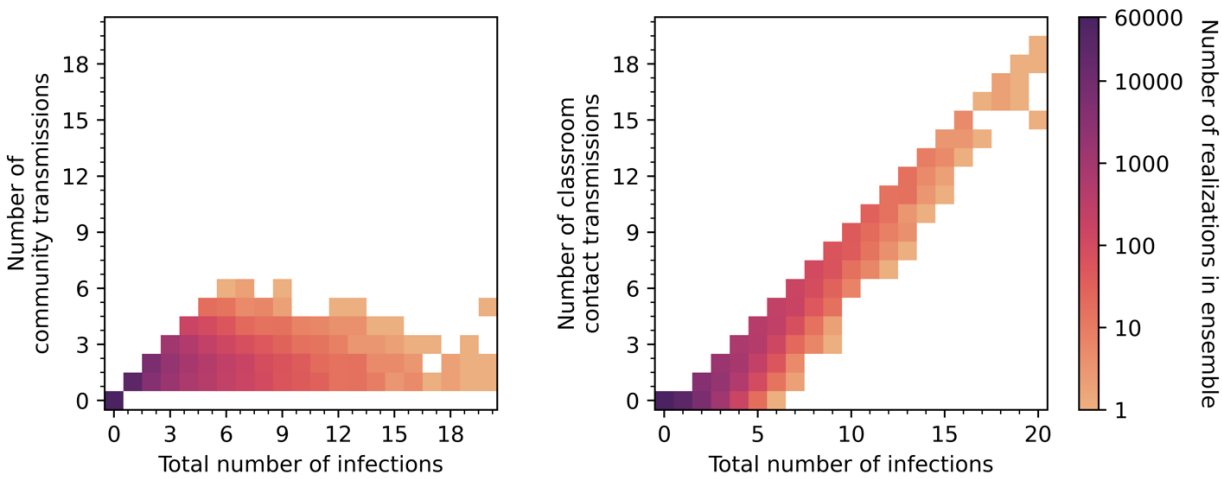

Figure 6. The number of infections caused by community transmissions and classroom contact transmissions in the ensemble simulations as functions of the total number of infections. The intensity of colour of the points indicates the frequency of occurrence within the ensemble on a logarithmic scale. Note, stochastic variations are apparent for low frequencies.

Assessing effectiveness of mitigation approaches

To examine the effectiveness of mitigation approaches, we perform stochastic ensemble simulations with different infection control measures imposed. We model a collection of 10 primary schools, each consisting of four classrooms with 30 pupils and one teacher. The Term 1 variation in community prevalence and incidence rate is adopted to model the changes over time in the external community epidemic and allows us to provide a comparison with the simulations above where there are no hard mitigation measures included.

In Figure 7 we compare the effect of the contact rates in classrooms and the mitigation approaches on the total number of pupils infected over the Term1 period within the 40 classrooms consisting of 1200 pupils in total. The box plots show the median, 25th and 75th percentiles, with the interquartile range (IQR) being their difference, the lower (25th percentile $- 1.5 \times$ IQR) and upper (75th percentile $+ 1.5 \times$ IQR) Tukey fences, and outliers.

We consider first the simulations where no 'hard' mitigation measures are applied (blue boxes in Figure 7). Here, the contacts rate distributions have a strong influence on the number of infections. With the contact rates of pre-Covid classrooms, the median number of pupils infected is 42 (75th percentile: 51). With reduced contacts, the median number of infections is reduced by 19% to 34, with decreases also in the spread of values, particularly on the upper tail (75th percentile: 41).

Each of the mitigation measures reduces the number of infected pupils, but they are generally less effective than the changes in the contact rates. Where contact rates are highest (i.e., using the pre-Covid contact rate distributions) the mitigation measures each substantially reduce the number of infections, as seen in the decrease in the quartiles, Tukey fences and outliers. The extent of the decrease in these statistics diminishes as contact rates decrease.

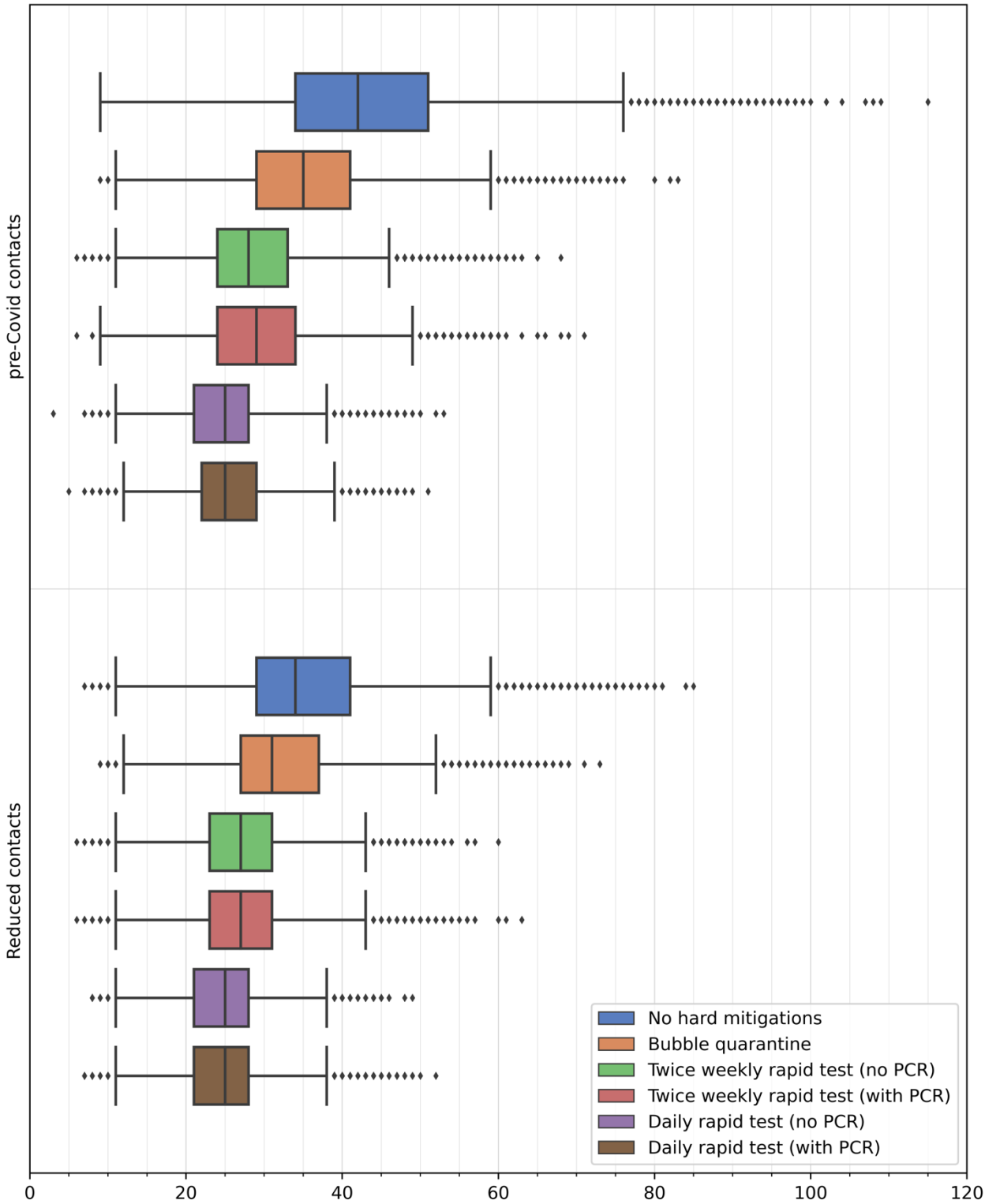

Figure 7. Comparison of the number of infected pupils in 40 primary classrooms with different mitigation measures applied. Box and whisker plots illustrate median, interquartile range, Tukey fences and outliers. Contact rate distributions for pre-Covid times and in small classes are derived from SEJ and used to construct distributions of expected contact rates for schools during Term 1 in September—October 2020. Mitigation measures simulated are: (i) 'No hard mitigations' where only individuals who are unwell are quarantined (blue); (ii) 'Bubble quarantine' where an unwell individual results in quarantine of the whole class (orange); (iii) 'Twice weekly rapid test (no PCR)' where each individual is tested with a rapid lateral flow test twice each week, with positive test results resulting in quarantine of the individual (green); (iv) 'Twice weekly rapid test (with PCR)' where each individual is tested with a rapid lateral flow test twice each week, with positive test results resulting in quarantine of the individual and a confirmatory PCR test is applied after a two-day interval (red); (v) 'Daily rapid test (no PCR)' where each individual is tested with a rapid lateral flow test daily, with positive test results resulting in quarantine of the individual (purple); (vi) 'Daily rapid test (with PCR)' where each individual is tested with a rapid lateral flow test daily, with positive test results resulting in quarantine of the individual and a confirmatory PCR test is applied after a two-day interval (brown).

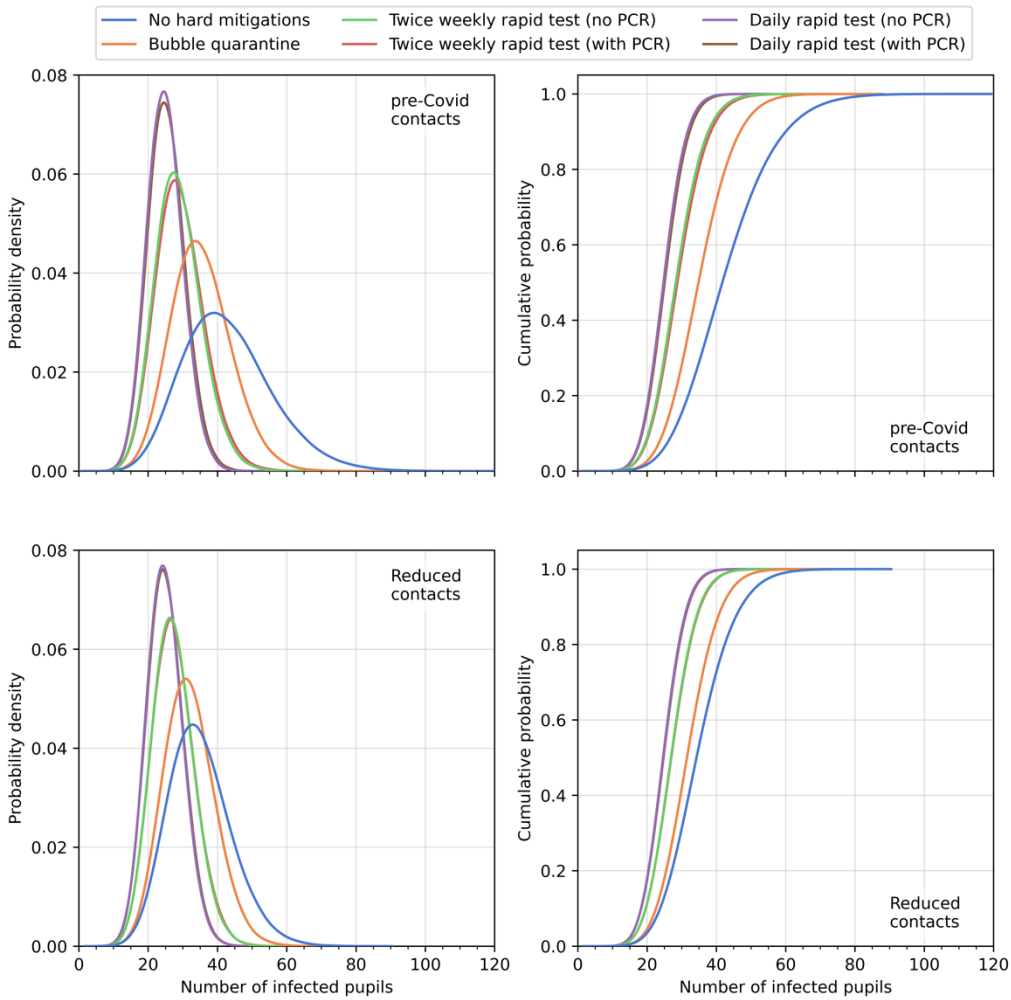

Figure 8. Comparison of the number of infected pupils in 40 primary classrooms with different mitigation measures applied. Estimates of the cumulative distribution functions (left) and probability density functions (right) using Kernel Density Estimation for each of the mitigation strategies are shown.

We compare each of the mitigation measures using the reduced contacts rates adopted for primary schools in Term 1 of 2020. Bubble quarantine lowers the upper quartile and upper fence values but does not substantially change the median number of infected pupils across the ensemble (Figure 7), lowering the median slightly from 34 to 31. This result is expected as bubble quarantine is not very effective in guarding against the seeding of infection carried into the classrooms from the community, and the spreading of this infection in the incubation period before an infected individual has symptoms that trigger the quarantine period. However, the reduction of the upper tail of the distribution (Figure 8) shows that bubble quarantine can produce a decrease in the occurrence of outbreaks. For example, small outbreaks of six of more infected pupils occur in 0.29% of isolated classroom ensemble realizations (reduced from 1% with no hard mitigations), outbreaks of ten or more occur in 0.01% of realizations (c.f. 0.1% with no hard mitigations), and large outbreaks of 15 or more infections do not occur in the ensemble (Table 2).

Considering next the use of rapid testing, we simulate approaches where each member of the classrooms is tested at regular intervals. Specifically, we model a twice weekly testing regime, such as that used in summer 2021, and a daily test regime. Further, we simulate each with and without confirmatory PCR testing, with a two-day time lag for conducting the PCR test.

With twice weekly testing without PCR testing, there is a decrease in the median number of infections to 27, a narrowing of the interquartile range with a pronounced reduction in the upper quartile to 31 infections, and a reduction in the upper fence value to 43 infections (Figure 7). This suggests that a test-based surveillance approach is effective in reducing infections within school through the disruption

of contact networks. Large outbreaks, in particular, are less common, occurring in approximately 0.01% of ensemble realizations (Table 2), and very large outbreak did not occur in the simulations. Confirmatory PCR testing has minor effects on the distribution of the number of infections, increasing only the maximum value in this sample to 63 (noting that the outliers are rare events and fluctuate when resampling). This result can be rationalized as there is a small non-zero probability of a 'false negative' PCR test overturning a 'true positive' rapid test and thus returning an infectious individual into the class. The frequency of outbreaks slightly increases in the ensemble.

Table 2. Proportion of realizations in ensemble of 100,000 isolated classroom simulations that produce outbreak with more than six infected pupils, large outbreaks with ten or more infected pupils, and very large outbreaks with 15 or more infected pupils. Six mitigation measures are simulated, as detailed in the main text. Zero value indicate no realizations occur in the ensemble. Simulations were conducted using the reduced contact rates distribution.

Mitigation measure	Proportion of outbreaks (>6 infected)	Proportion of large outbreaks (≥ 10 infected)	Proportion of very large outbreaks (≥ 15 infected)
No hard mitigations	1.06%	0.27%	0.023%
Bubble quarantine	0.29%	0.03%	0
Twice weekly rapid test (no PCR)	0.09%	0.009%	0
Twice weekly rapid test (with PCR)	0.13%	0.013%	0
Daily rapid test (no PCR)	0.003%	0	0
Daily rapid test (with PCR)	0.014%	0	0

Implementing daily rapid lateral flow testing frequency further shifts the distribution of the number of infections to lower values (Figure 7 and Figure 8), although the change is modest (median: 25; 75th-percentile: 28 for simulations both with and without PCR testing), representing a change of only 7% in the median and 10% in the upper quartile values compared to the twice-weekly rapid testing ensembles. Outbreaks are further reduced (albeit from small values for twice-weekly testing), and no large outbreaks of ten or more infected pupils occur in the ensemble (Table 2). In this case the effect of confirmatory PCR testing is barely discernible in the numbers of infected pupils, with only far outliers differing (the interpretation of which should be made cautiously without larger ensembles). This can also be explained: daily testing ensures that 'false negative' test results are rapidly re-tested, and sequences of repeated false negatives are unlikely even for the low sensitivity rapid test. There is an increase in the number of outbreaks within the ensemble, although the proportion remains small.

Where infection occurs in the ensemble realizations, we track the number of secondary infections within the classroom from each infected individual over the duration of the simulation. The mitigation measures shrink the number of secondary infections by reducing contact transmission. This is illustrated in Figure 9 which shows two summary statistics (the mean and 99th-percentile values) to characterize the distribution of the number of secondary infections that occur in the ensemble simulations. To highlight the trends when curtailing contact rates, we have conducted simulations here with the contact rate distributions derived from SEJ for small classes and applied these rates to full classroom attendance; it is unlikely that such low contact rates could be maintained in a fully occupied classroom, but the simulations are illustrative hypothetical scenarios.

The mean value of the number of secondary infections is less than unity in all cases, indicating that the infection is not transmitted in the classroom contacts of most primary cases. Each mitigation measure reduces the mean value of the number of secondary infections, with more pronounced effects when the contact rate is high.

Outbreaks in the classroom can be driven by relatively rare '*super-spreaders*'. Here our model includes the influence of having some very gregarious individuals in a mixing group, as reported in the contact

in [1]. Here we illustrate the impact of mitigation measures on super-spreaders by calculating the 99th percentile value for the number of secondary infections (i.e., 1% of infected individuals in the simulations have a secondary infection number larger than the 99th percentile value). The mitigation measures are again seen to reduce the 99th percentile values, although the 'bubble quarantine' approach has only modest (if any) affect. The testing-based surveillance approaches are more effective in eliminating super-spreading, reducing the 99th percentile value of the number of secondary infections to unity (or smaller) if daily testing is implemented.

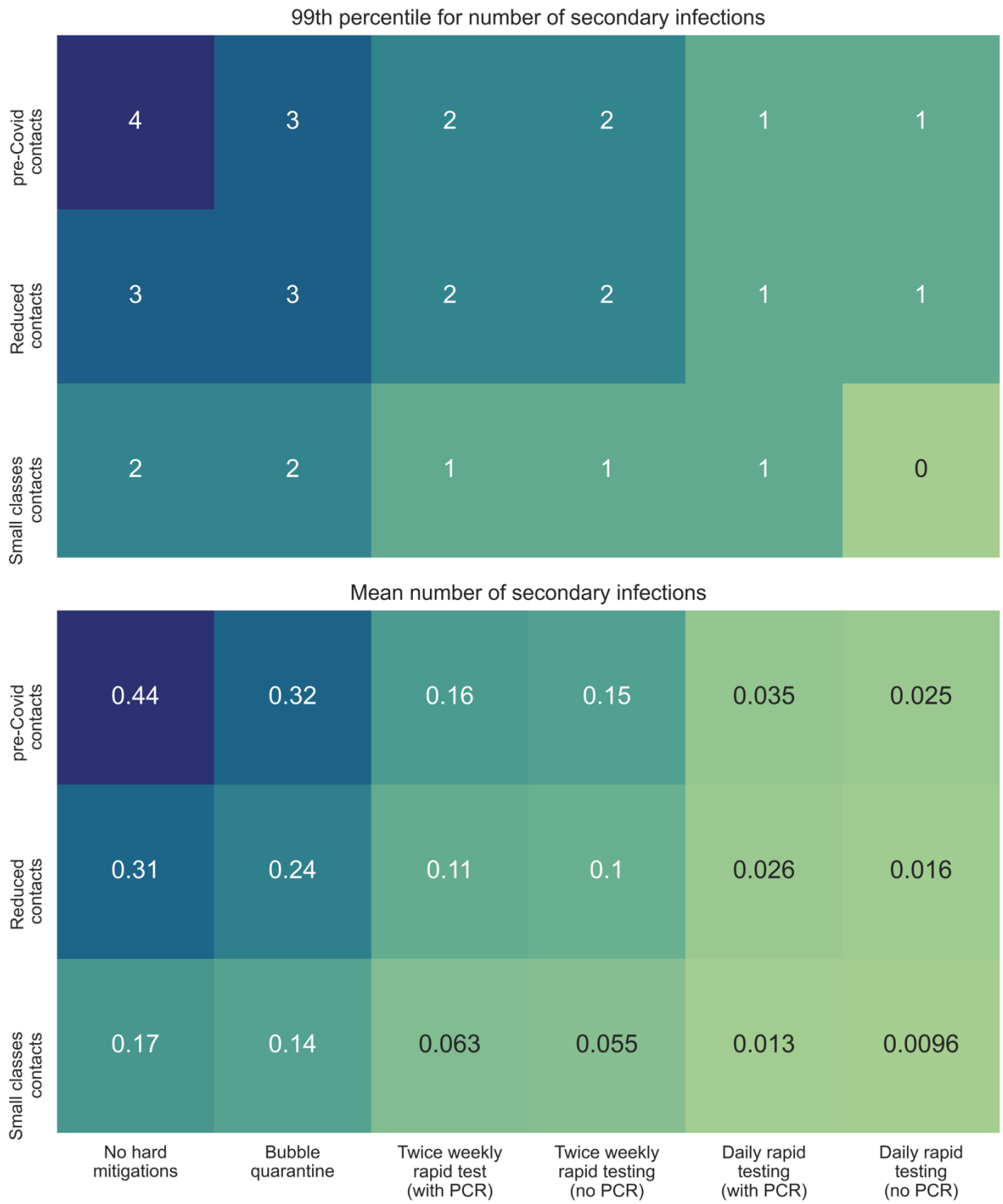

Figure 9. The number of secondary infections in classroom simulations with different contact rates distributions and mitigation measures applied. Two summary statistics to characterize the distributions are shown, with the 99th percentile value (upper panel) and mean value (lower panel) of the number of secondary infections in all 100,000 ensemble simulations calculated. 'Small classes contacts' refers to simulations that adopt contact rate distributions derived from SEJ for small classes that are applied to full classroom attendance. Note, most of infections in the simulations are not transmitted in classroom contacts, so mean values are below 1.0 in all cases

The mitigation measures have consequences on pupil absences. The 'bubble quarantine' is particularly severe, with whole classes being quarantined for a period of 10 days if there is a single symptomatic case. However, the testing-based surveillance approaches leads to self-isolation of individual cases, so the impact on absences is greatly reduced. In Figure 10 we illustrate the effects of the mitigation measures on the number of pupils present in the classroom on one day, 23 October 2020 which we took as the last day of Term 1 in 2020. As the majority of the realizations in the ensemble have no infections, there is typically full attendance in each of the scenarios simulated. However, when 'bubble quarantine' is employed, there is full-class quarantine on this day in greater than 15% of the ensemble members. If this is result is applied to all classrooms in England, this would suggest approximately 25,000 classrooms in quarantine on this day and, with an average class size of 27 in England [23], approximately 675,000 pupils absent from school. For the testing-based surveillance, there are fewer absences, with only ~0.1% of the ensemble simulations having more than five pupils in quarantine on the last day of the school term. The contact rate distribution is again seen to play a prominent role, with reduced contacts leading to reduced absences as there are fewer infections to trigger quarantine.

To assess the impact of the testing strategies, we focus here on the 'Reduced contacts' distribution (lower panel in Figure 10). Twice weekly rapid testing has similar proportions absent from the classroom as the simulations with no hard mitigations applied. Without confirmatory PCR testing, twice weekly rapid testing slightly increases the number of ensemble members for which there is a single pupil absent, while the application of a confirmatory PCR test returns this proportion to close to that found when no hard mitigations measures are applied. In contrast, daily testing without PCR confirmation has a substantially elevated proportion of simulations with classroom absences, with more than ~1% of simulations having five pupils absent. Confirmatory PCR testing reduces the absentee proportions to below that when no mitigation measures are applied. This confirms that many of those quarantined by daily rapid testing are done so unnecessarily and are promptly returned to the classroom when a PCR test overturns a false-positive rapid test. However, the disruption to education by missing some days at school is significant and the benefit from a public health perspective is negligible.

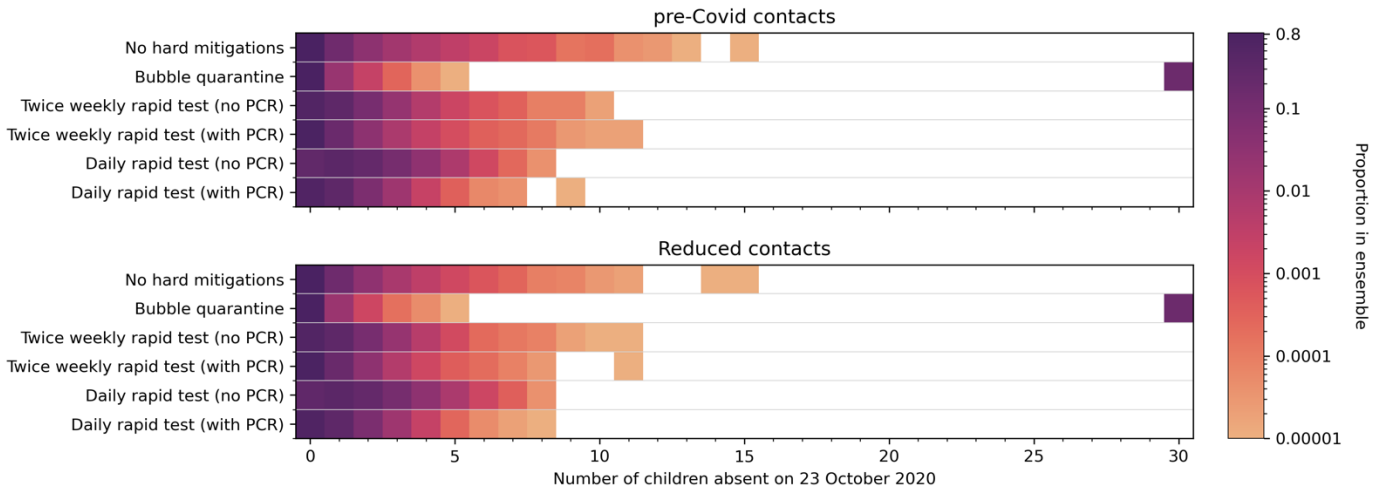

Figure 10. Number of pupils absent from school on 23 October 2020 (the last day of Term 1 in 2020) in an ensemble of 100,000 simulations of a class of 30 pupils. Two contact rates are applied: pre-Covid contact rates (top panel) and reduced contact rates (bottom panel). Six mitigation measures are modelled. Colours indicate the proportion of ensemble realizations on a logarithmic scale (white indicating no occurrence in the ensemble).

We estimate the numbers absent on a national scale by performing a bootstrap resampling of our ensemble to produce a sample of 156,843 classes, matching the number of state primary schools in 2020—2021 [23]. As our nominal class size is 30 pupils, this gives a total number of pupils of 4,705,290 which is comparable to the national state-primary pupil count of 4,177,058 [23] (a different of -13% in the simulated classes). The bootstrap ensemble is recomputed 10,000 times to construct and median value for the number of pupils absent on 23 October 2020 and confidence intervals on our estimates, which are reported in Table 3.

Table 3. Estimated national number of pupils absent from school on 23 October 2020 from ensemble simulations with bootstrap resampling. Isolated classes are simulated with six mitigation strategies and two contact rate distributions. Ensembles of 100,000 classes are simulated and collections of 156,843 classes are formed by bootstrap resampling from the ensembles. Each bootstrap is constructed 10,000 times to obtain distributions on the number of absences, and we report the median, 5th and 95th percentile values. The percentage of the school pupil population is estimated assuming each class has 30 pupils, giving a school population of 4,705,290.

		Number of pupils absent		**Percentage of pupils absent**	
	Mitigation measure	median	5th; 95th percentiles	median	5th; 95th percentiles
Pre-Covid contacts	No hard mitigation	48,384	47,861; 48,908	1.03	1.02; 1.04
	Bubble quarantine	762,605	755,422; 769, 729	16.21	16.05; 16.35
	Twice weekly rapid test (no PCR)	95,529	94,977; 96,069	2.03	2.02; 2.04
	Twice weekly rapid test (with PCR)	45995	45,592; 46,416	0.98	0.97; 0.99
	Daily rapid test (no PCR)	202,887	202,152; 203,617	4.31	4.30; 4.33
	Daily rapid test (with PCR)	81,329	80855; 81,811	1.73	1.72; 1.74
Reduced contacts	No hard mitigation	40,734	40,312; 41,174	0.87	0.86, 0.88
	Bubble quarantine	748,469	741,461; 755,625	15.9	15.8, 16.1
	Twice weekly rapid test (no PCR)	93,517	92,980; 94,053	1.99	1.98, 2.00
	Twice weekly rapid test (with PCR)	42,991	42,615; 43,374	0.91	0.91, 0.92
	Daily rapid test (no PCR)	203,566	202,836; 204,305	4.35	4.31, 4.34
	Daily rapid test (with PCR)	80,725	80,248; 81,199	1.72	1.71, 1.73

Table 3 shows that the different mitigation strategies have a strong effect on school absence, with the model adopting bubble quarantine resulting in more than 15% of pupils absent from school on 23 October 2020. The using of rapid testing reduces the numbers absent substantially, decreasing the number of absent pupils to around 2% of the national primary school pupil population for twice weekly testing with no PCR confirmation. This is further decreased to around 1% of the pupil population if confirmatory PCR testing is used to return pupils with false-positive rapid test results to the classroom. However, daily rapid testing produces many more false-positive results, so that the number of absent pupils increased markedly to over 200,000 (approximately 4% of the pupil population) when there is no confirmatory follow-up PCR test. With a PCR test to identify false-positive cases of daily rapid tests, the number of absent pupils decreases to approximately 80,000 (1.7% of the pupil population) which is comparable to the number absent with twice-weekly rapid testing with no PCR follow-up.

5. Forecasting SARS-CoV-2 in schools in 2021

Amendments to the model

To apply our model to forecasting SARS-CoV-2 infection transmission and effectiveness of mitigation measures to the first seven weeks of the new school year (i.e., assumed here to be Thursday 2 September to Friday 22 October 2021 of Term 1), we alter our model parameter from those used for Term 1 2020. In particular, we include the impact of population vaccination by September 2021. Here we assume that there is 95% uptake amongst classroom staff, and that 1% of children have been vaccinated.

Another crucial difference is the high prevalence of the Delta (B.1.617.2) variant in the UK population. Genomic surveillance [28] and modelling studies [15-17] indicate that the Delta variant is expected to be predominant in the UK in September 2021, accounting for more than 95% of Covid-19 infection cases [15-17].

The Delta variant has two major impacts on our updated model. Firstly, it has substantial transmission advantage over the original wildtype SARS-CoV-2 [29, 30]. This is estimated as 61% transmission advantage over the Alpha (B.1.1.7) variant in [29], from which we estimate as a transmission advantage of 140% over the original wildtype. We therefore increase our values of the probability of contact transmission from original wildtype SARS-CoV-2 values using this scaling factor. The probabilities

of contact transmission we adopt for the Delta variant are given in Table 4 (again assuming, as before, that adults are twice as likely as children to transmit infection [5-7]).

Table 4. Probability of transmission of Delta variant SARS-CoV-2.

Delta-variant transmission probability	Adults	Children
Symptomatic	0.084	0.042
Asymptomatic	0.059	0.029

An additional characteristic of the Delta variant, which drives increased infection, is a shortening of the incubation period for infected individuals [31]. Delta-variant infection is detectable (by PCR testing) in cases approximately two-days earlier than for original wildtype SARS-CoV-2 infection [31]. We adopt this shorter interval in our model by reducing the median duration of the incubation period by two days. Thus, for each individual in our population, we assign an incubation time drawn from log-Normal distribution [4] with $t_i \sim \log N(1.10, 0.5)$. In principle, the variance of the log-Normal distribution could also be amended, but in the absence of detailed information, we retain the previous value. The reduced incubation period affects the time-dependence of the relative transmission probability, with a more rapid increase to peak-infectiousness for the Delta-variant model parameters. Therefore, the probability of infection transmission from pre-symptomatic individuals is enhanced when considering the Delta-variant.

Forecast scenarios

There is significant uncertainty in the future trajectory of the epidemic into the new school year. Projections from epidemiological models fitted to observations [15-17] provide useful indicators on which we can base forecast scenarios. Typically, the projections in [15-17] suggest a third wave of the epidemic is likely in late-summer as the effects of easing of restrictions on 19 July on infection transmission are realized. The model projections also indicate that infection amongst younger age-groups becomes dominant [15]. The magnitude of a third wave varies across models and with model assumptions, particularly regarding social mixing over the summer, and peak incidence rates (per 10,000 population per day) range from ~12 [15] to ~300 [16] across the models, spanning values substantially higher than those for September 2020 (Figure 4; [14]).

The timing of the peak incidence rate in any third wave is also highly uncertain and sensitive to modelling assumptions, with many model results projecting the peak will pass during August 2021, while others suggest peaking in mid-September [15, 16]. Thus, the start of school year 2021—2022 may well coincide with a time of elevated, albeit falling, incidence rates, but a few weeks into term-time could see peaking prevalence and still increasing incidence of infection. The dynamics of Delta variant infection, peak timing of any third wave and its magnitude will all affect levels of infection in classrooms. These factors can be incorporated into our model predictions by altering the seeding of infection based on community prevalence at model initiation and on the community incidence rate through community transmission.

To model the community incidence rate of new infections over Term 1 of 2021, we digitize projection graphics produced by a national scale epidemiological model in [15]. This provides two time series of incidence rate which differ in the magnitude and timing of the peak depending on how rapidly there is a return to pre-Covid contact levels (Figure 11) which we refer to as the 'gradual-relaxation' and 'rapid-relaxation' incidence rates. For the gradual-relaxation scenario, a peak incidence rate of approximately 67,000 new infections per day was projected to occur in late July, with a gradual decrease through August and September, to approximately 20,000 per day on 1 October 2021, with incidence slightly increasing subsequently [15]. In this scenario, the school term begins in a period of relatively steady incidence. In contrast, if societal contact rates rapidly return to pre-pandemic levels in summer 2021, then the projected peak incidence rate was approximately 145,000 new infections per day and occurs on 12 August 2021, with incidence rates elevated with respect to the gradual-relaxation until 1 October

2021, from when the incidence rate gradually decreases to low levels through October and November [15]. The incidence of new infection was not expected to be evenly distributed through the population, but rather skewed to lower age groups who were unvaccinated. Here, we adopt the values reported in [15] that attribute 27% and 25% of new infections to age groups 0—9 and 10—19, respectively, for the gradual-relaxation scenario, and combine these to produce a child-incidence rate. These proportions change but slightly for the rapid-relaxation scenario to 26% and 24% for age groups 0—9 and 10—19, respectively.

There is currently limited data to compare these projections, and we recognize that the projections are sensitive to changes in model parameters and external forcings (e.g., weather conditions influencing the extent of indoor mixing in the summer). Using data from the Office for National Statistics Coronavirus Infection Survey [24] from June 2021 until the latest available, the gradual-relaxation scenario appears to better track the incidence rate. However, the traditional summer holiday period in August is included in the ONS incidence data at the time of writing (25 August 2021) and the UK Coronavirus Dashboard (https://coronavirus.data.gov.uk/) reports a nationally aggregated incidence rate of 224,316 per week (as a "*7-day rolling rate by date of specimen as of 19 August 2021*"), based on laboratory PCR test or self-reported lateral flow tests. Assuming this weekly incidence rate is equally distributed over each day, this corresponds to an approximate daily rate of approximately 32,000 per day. This aligns with the model projections for the gradual-relaxation scenario [15]. However, these data sets have a trend of increasing incidence rate since 6 August 2021. Thus, neither of the projection scenarios captures all aspects of the currently available data. However, they represent a useful basis for assessing possible classroom infections as schools open in September 2021.

Our model also requires the community prevalence. This is not reported in the projections, so we estimate prevalence using a simple rescaling of incidence rates with comparison to the available data (Figure 11), as described in the supplementary material. The prevalence has less importance in our model than the incidence rate (as it is used as a seeding of infection on initiation and in cases where substitute teachers are added to the simulated classroom) and we judge that the estimate is sufficient in our forecasts.

Guidance from the Department for Education issued in summer 2021 [32] has removed many of the SARS-CoV-2 mitigation measures in school settings, but recommends 'soft mitigation' measures continue to be applied, such as enhanced cleaning and classroom ventilation. Therefore, in our forecast simulations, we do not model any of the possible hard mitigations that might be activated under certain conditions. Additionally, schools will be able to reconvene in large groups, such as during lunch and break times and for school assemblies. This may increase transmission rates within school settings and particularly between classes. However, school managers are responsible for deciding whether to continue to 'bubble' classes to minimize interactions. Teachers and school leaders may also decide to continue to implement measures that reduce contacts within the classroom, so we adopt the two contact rates distributions from [1] that model contact rates for pre-Covid classrooms and reduced contact rates during the epidemic.

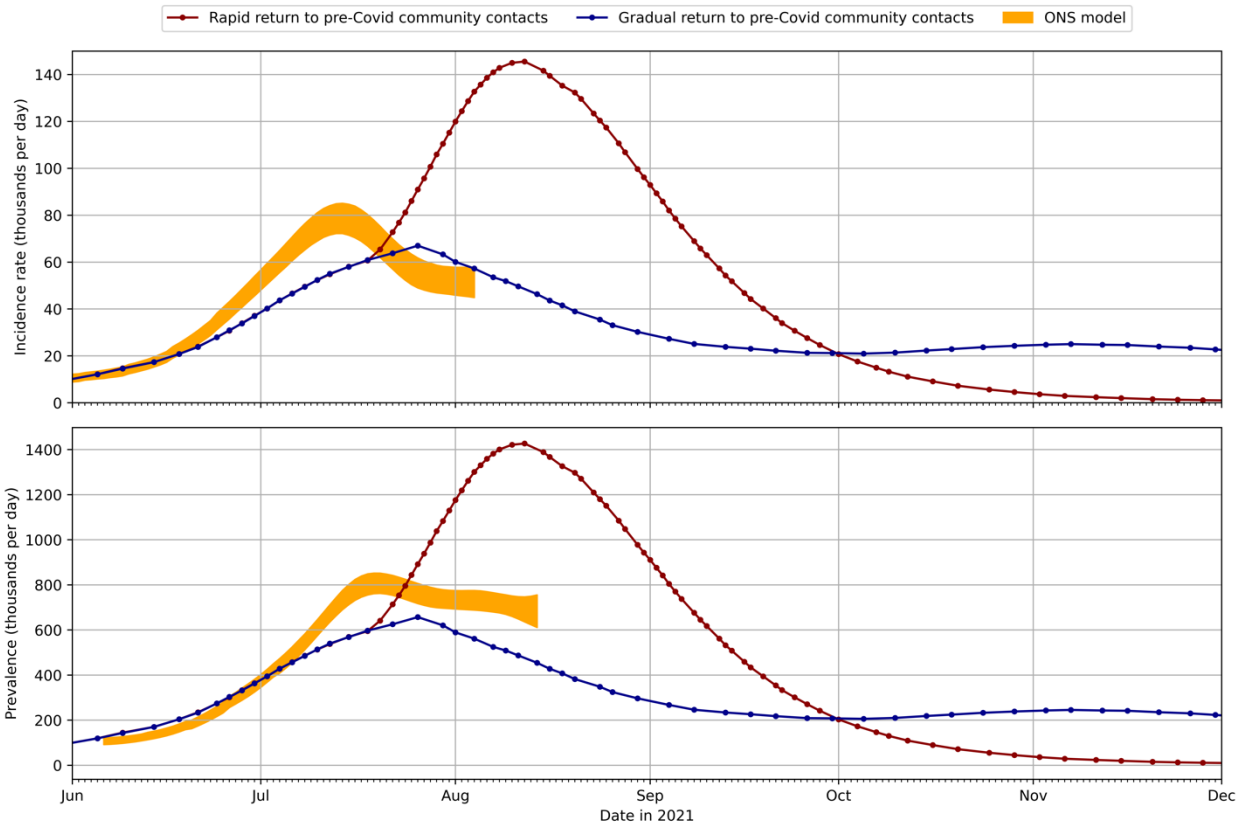

Figure 11. Projections of the incidence rate (upper panel) and prevalence (lower panel) in the UK from June to December 2021. Incidence rates are obtained from a national scale epidemiological model [15] with scenarios modelling a gradual or rapid return to pre-pandemic societal contact rates. The prevalence is modelled assuming $R_0 = 5$ and infection duration of 40 days (see supplementary material). Data from the Office for National Statistics Coronavirus Infection Survey [24] for the modelled incidence and prevalence with credible intervals are illustrated for the data available at the time of writing.

Results

We perform stochastic simulations for an isolated school classroom over the 50 days of Term 1 in 2021, adopting the 'reduced contacts' distribution of contact rates (i.e., we assume that in this time period, schools will continue to enact measures to reduce the number of contacts in classrooms).

Figure 12 illustrates the result when for the 'gradual-relaxation' scenario, showing time series of the number of infected pupils in each realization from an ensemble of 100,000 stochastic simulations in which no hard mitigation measures are applied. Therefore, Figure 12 is directly comparable to Figure 5, and shows there is potential for substantially increased infection within school classrooms in Term 1 of 2021 when compared with Term 1 of 2020. We find both (i) a potential substantial increase in the proportion of ensemble realizations that have infected pupils (62% for 2021, compared to 46% for 2020) and (ii) the maximum number of simultaneously infected pupils in a realization could be increased (22 in 2021, compared to 18 in 2020). There are also more realizations in the ensemble that have several simultaneously infected pupils, although most frequently there is only a single infected pupils in the class, as we found in the Term 1 2020 simulations.

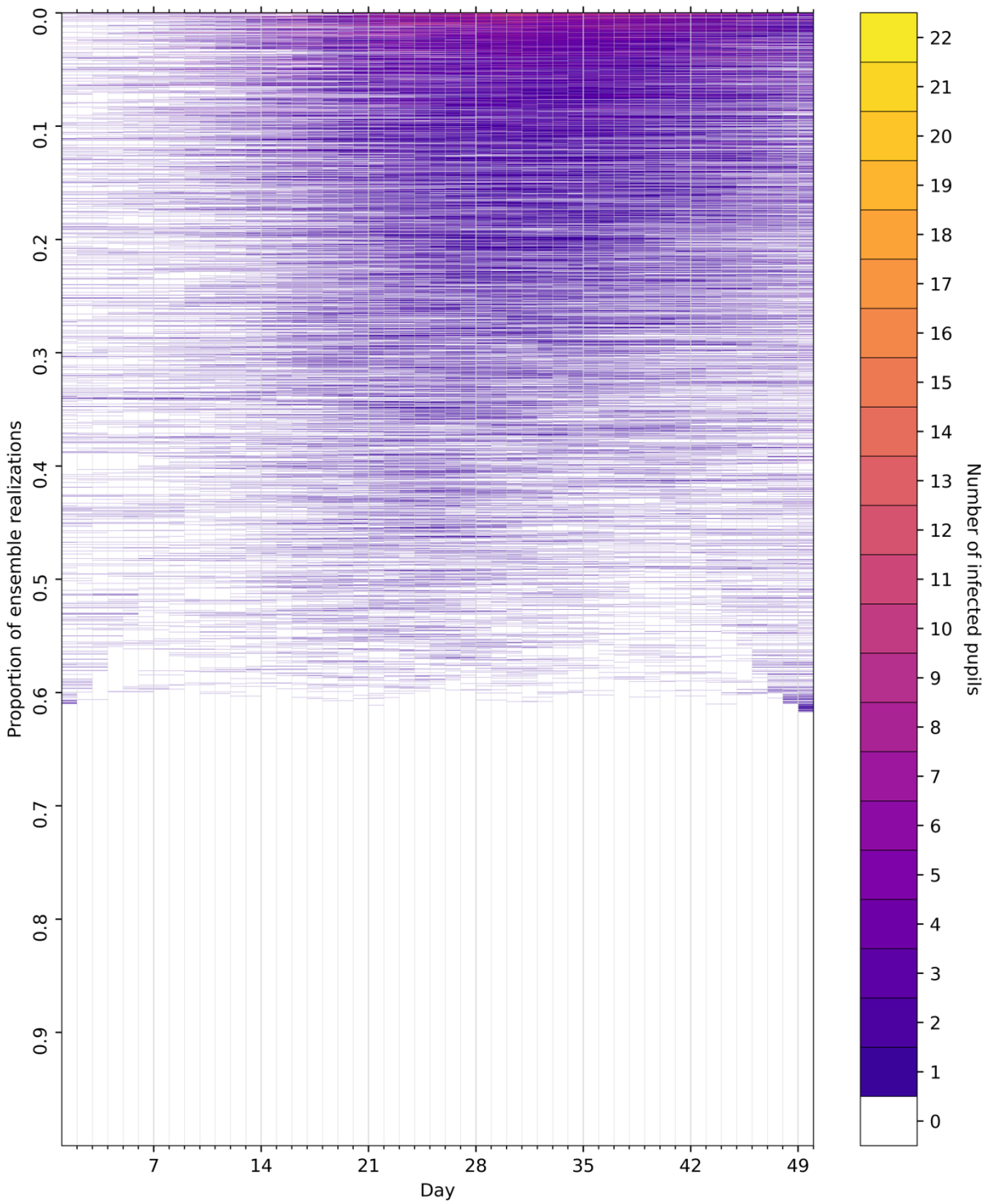

Figure 12. Time series of the number of infected pupils for each realization in an ensemble of 100,000 stochastic simulations of an isolated classroom in Term 1 of 2021 under the gradual-relaxation scenario. Each row of the heatmap corresponds to an ensemble member, with colours denoting the number of pupils in the classroom with Covid-19 infection on each day. The ensemble realizations have been sorted by the sum of the number of infected pupils over all days. Colours denote the number of infected pupils in the classroom for each day of the simulation.

The gradual-relaxation scenario can also be compared to forecasts under the rapid-relaxation scenario (Figure 13) where the incidence rate is approximately three times larger at the start of Term 1. In this case, there are infected pupils in more than 89% of the ensemble realizations, and more realizations have substantial numbers of pupils simultaneously infected. This is confirmed in Figure 14 where histograms (empirical pdfs) of the total number of infected pupils in each realization are shown for both scenarios for Term 1 2021 and the equivalent simulation for Term 1 2020.

The results show that for both scenarios for Term 1 of 2021, there is a probability of greater than 0.01 of classes having 10 or more infected pupils over the course of the term. Indeed, the number of infected pupils at the 1% probability level increases from 5 for the 2020 simulation, to 7 for the gradual-relaxation scenario for 2021, and to 13 under the rapid-relaxation scenario. The change in the upper tail values from Term 1 2020 (with no hard mitigations) to the 2021 scenarios is even more pronounced: the 99.9 percentile values on the numbers of infected pupils are 12 for 2020 and 22 for the 2021 scenarios. Thus, the simulations indicate a much greater potential for substantial outbreaks in Term 1 of 2021.

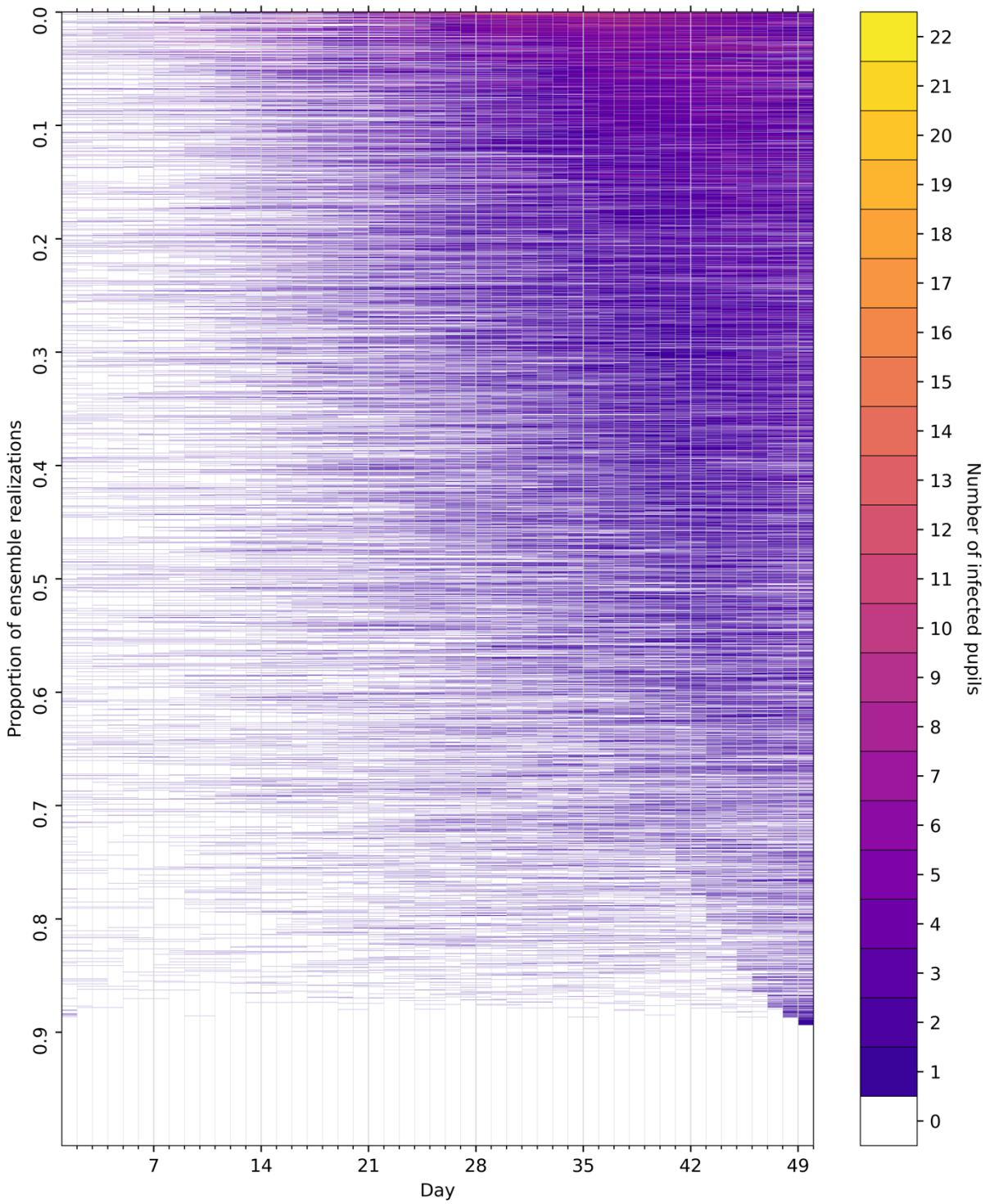

Figure 13. Time series of the number of infected pupils for each realization in an ensemble of 100,000 stochastic simulations of an isolated classroom in Term 1 of 2021 under the rapid-relaxation scenario. Each row of the heatmap corresponds to an ensemble member, with colours denoting the number of pupils in the classroom with Covid-19 infection on each day. The ensemble realizations have been sorted by the sum of the number of infected pupils over all days. Colours denote the number of infected pupils in the classroom for each day of the simulation.

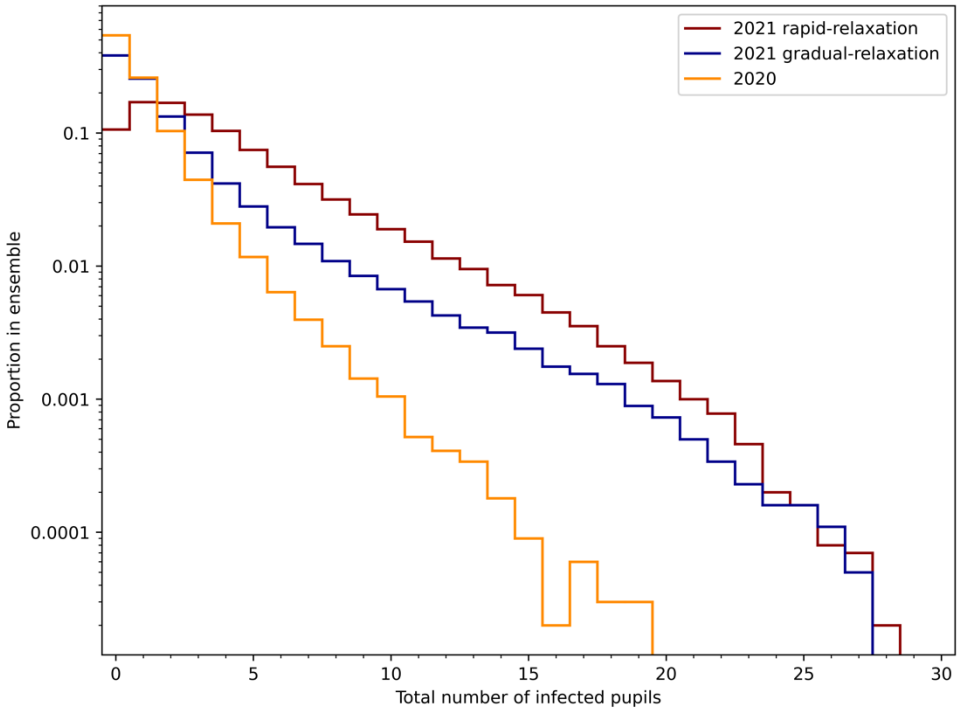

Figure 14. Histograms (empirical pdfs) of the total number of infected pupils in ensembles of 100,000 simulation of the rapid-relaxation and, gradual-relaxation scenarios for Term 1 in 2021 and the 'no hard mitigations' simulation with reduced contacts for Term 1 in 2020. Note the logarithmic scale on the proportion in the ensemble.

Simulations have also been performed for the Term 1 of 2021 scenarios assuming the classroom contacts rates return to pre-Covid levels. To compare these simulations with those above, we consider again the collection of ten schools each with four classrooms of 30 pupils. This better illustrates the level of infection that local authorities, city boroughs or academy networks may need to respond to in Term 1 of 2021.

In Figure 15 we illustrate the differences that may occur in these scenarios by plotting histograms that show the probability that a given number of classes in the collection experiences infection at different levels. We find that the probabilities of classes having no infections (Figure 15a) is independent of the classroom contact rates, but strongly depends on the assumed societal contact relaxation scenario, with a rapid-relaxation to pre-Covid societal contacts substantially reducing the number of classes with no infections. Similarly, the number of classes experiencing a single infection (Figure 15b) has only slight variations with the classroom contact rates, but differ substantially under the two societal contacts forecast scenarios.

The effects of in-classroom contact rates increase if we consider outbreaks. Small clusters of three or more infections (Figure 15c), which would be likely to trigger 'extra action' under current DfE guidelines [26], are more likely for high (pre-Covid) in-class contact rates of than for reduced contacts. However, differences in the community incidence rate in the two forecast scenarios have the greater effect, three of more infections likely in around half of the 40 classes in the collection under the rapid-relaxation scenario, but in only 5—8 classes for the gradual-relaxation scenario.

For outbreaks of more than six infections in a class (Figure 15d), our model suggests 16 classes in the collection are likely to experience such infection levels with pre-Covid classroom contact rates under the rapid-relaxation scenario, reducing to 12 classes under the gradual-relaxation scenario. Outbreaks of ten or more infections (Figure 15e) are less likely in all simulated conditions, but in the worst-case modelled (pre-Covid classroom contact rates under a rapid-relaxation of societal contacts) it remains likely that 8 classes in the collection would experience an outbreak of this size. Additionally, the number of classes with ten or more infections is similar for reduced classroom contacts in the rapid-relaxation scenario and the pre-Covid contacts in the gradual-relaxation scenario.

Finally, considering large outbreak for 15 or more pupils (i.e., at least half of the pupils being infected), we find that this is likely in three classes, and may occur in more than ten classes, under the worst case modelled, but becomes unlikely in the gradual-relaxation scenario. However, even in the best case modelled (reduced classroom contact rates and gradual-relaxation of societal contacts) there is an appreciable probability (approximately 0.0075) of three classes experiencing a large outbreak.

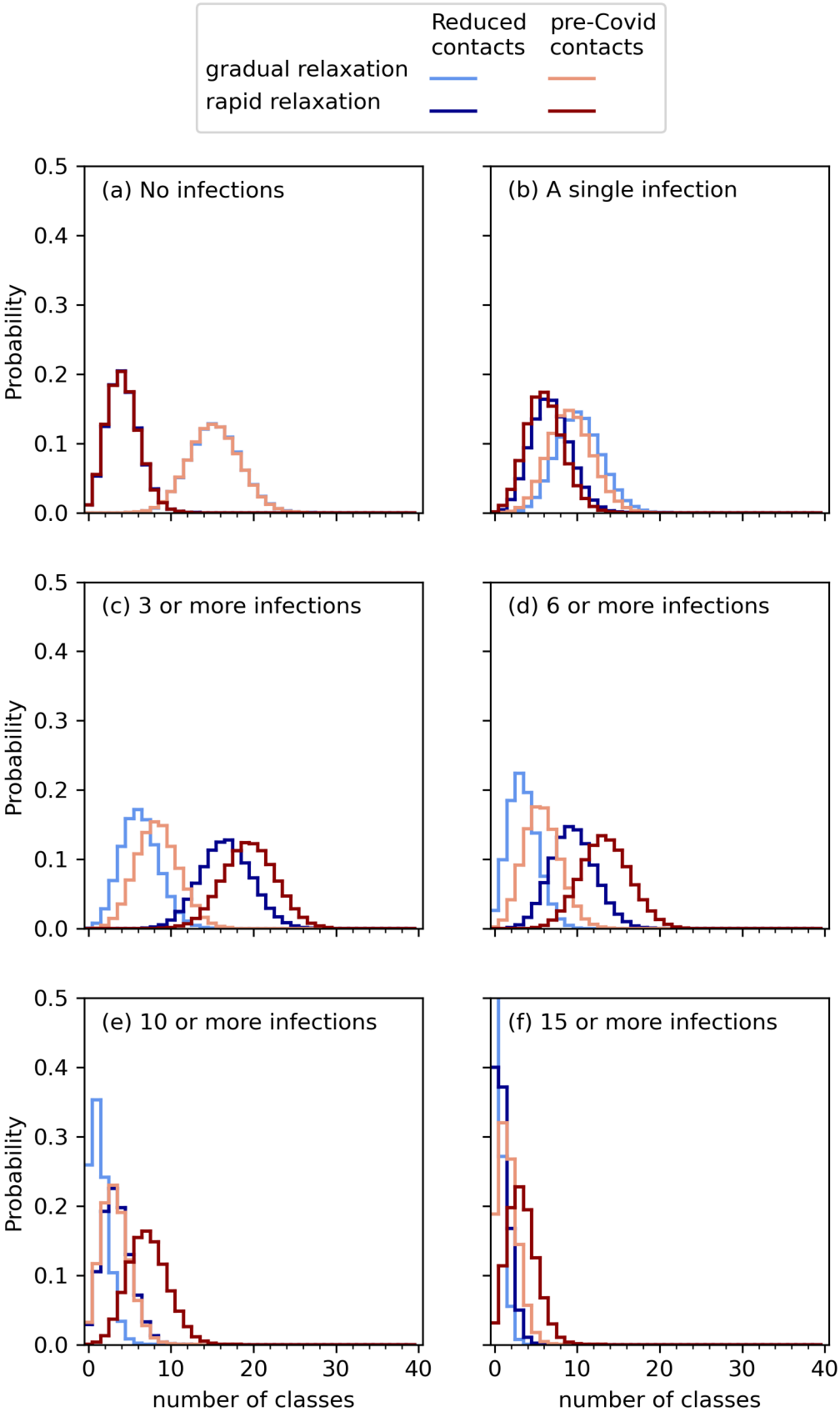

Figure 15. Histograms of the number of classes in a collection of ten schools each with four classes of 30 pupils that are forecast to experience SARS-CoV-2 infections in Term 1 of 2021 under different scenarios: gradual- and rapid-relaxation to pre-Covid societal contacts, and reduced or pre-Covid contact rates in the classroom. Six different infection levels are considered: (a) classes with no infections; (b) classes with a single infection; (c) classes with at least one infection; (d) classes with more than six infections; (e) classes with ten or more infections; and (f) classes with 15 or more infections. Note, in (f) the reduced contacts gradual-relaxation scenario has a probability of no classes with 15 or more infections of 0.66 and is curtailed in these plots to better illustrate the lower probability by high consequence outbreaks

6. Discussion

Main results

We have developed a stochastic discrete compartmental model of SARS-CoV-2 transmission in primary school classrooms and applied this to analyse infections over Term 1 of 2020, and to forecast Term 1 of 2021. The model takes account of the known large variations in the contact patterns of individual pupils in primary schools [1]. This includes allowing for the possibility of having one or more very gregarious individuals in a mixing group, who may drive major in-class outbreaks, and allows the potential impact of such pupils to be investigated quantitatively. Our basic model is for transmission within a class, but individual schools can be modelled as collections of such classes such that the results can be applied to local authority areas and academy networks and provide estimates on a national scale.

Our analysis of model findings for Term 1 of 2020 has focussed on assessing alternative mitigation measures that have been proposed to manage infection risk in schools. While there are limited data with which to test our model, the reanalysis broadly aligns quantitatively with observations from 2020. Furthermore, in comparing mitigation measures, our model parameters are not varied, allowing relative changes to be compared and reducing the impact of uncertainties on parameters.

We find the dynamics of infection in the classroom follow the external community incidence rates through the seeding of infection. Typically, there are few instances of classroom transmission and, in the majority of ensemble realizations, there are either no infections or a single pupil is infected.

Our model results support the notion that outbreaks can be defined as infections exceeding about 5 persons in a class-sized mixing group. Below 5 cases, we find that the infections are largely driven by community prevalence, but this connection is broken for realizations numbering more than 5 infections with large outbreaks driven by classroom transmission. Our forecasts indicate that outbreaks may be substantially larger and more frequent in Term 1 of 2021 compared to Term 1 of 2020. This is due to the higher transmission rates with variant Delta SARS-CoV-2 and its more rapid onset of infectivity. In the worst-case scenario, the modelling indicates that almost all school classes will experience at least one infection case in Term 1 of 2021.

Our comparison of mitigation measures over Term 1 of 2020 indicated that bubble quarantine is ineffective in reducing infection in schools when compared to testing-based surveillance approaches. This is due to the predominance of small numbers of infected pupils and reduced interactions and transmission across within the in-class contact network at that time. In such cases, bubble quarantine places pupils into quarantine and disrupts education to achieve only modest reduction in the probability of infection outbreaks. In contrast, the testing-based surveillance approaches achieve a greater reduction in numbers of infected pupils with much less disruption through isolation away from the classroom, very substantially reducing probability of pupil absences. The frequency of testing has a small effect on the upper tail of the number of infections, but frequent testing is likely to be burdensome, particularly for young children who cannot self-administer tests. Confirmatory PCR testing has very little effect on the infection levels in the classroom but can reduce the time of quarantine out of the classroom for false-positive rapid tests.

A prominent result of our modelling is the essential role of contact rates within the classroom on infection transmission. In this study, we have compared two different contact rate distributions, derived from an elicitation of teachers in May 2020 [1], using our new agent-based stochastic transmission model. The reduction of contact rates from pre-Covid levels is shown to have a comparable reduction on pupils' infection as implementing bubble quarantine with pre-Covid contacts. The reduction of contacts also shrinks in-class secondary infections to a similar extent as the application of bubble quarantine.

Following on from these retrospective analyses, we have applied the new model to forecasting infection in schools for Term 1 2021. Here the model assumptions and parameters have been changed from the September 2020 situation to include a largely vaccinated adult population and the more virulent Delta variant of SARS-CoV-2. We have contrasted two different contact patterns, namely pre-Covid with no mitigation and then reduced contact levels, as ascribed to primary schools in Term 1 of 2020. We then contrasted two different modelling projections, concerning community prevalence and incidence rate of infection, making different assumptions about how quickly the public would alter their community contact patterns back to pre-Covid rates. In comparing these two incipient scenarios for September--October 2021, we conclude, commensurate with [5], that prevalence levels in the community will remain the dominant control on infection in primary schools.

Adopting 'soft' mitigation measures to reduce transmission has a modest effect compared to differences in projected community prevalence but significantly suppresses occurrence of larger outbreaks.

Implications for managing Covid-19 in schools

Our results have implications for the management of SARS-CoV-2 infection transmission in primary schools. We find that mitigation steps that reduce contacts between pupils have a tangible benefit in disrupting transmission. Schools are already experienced in organising classrooms and other social interactions to reduce pupil-to-pupil and pupil-to-teacher contacts. Our model results suggest that, given practicalities of managing contacts within schools, continuation of these practices can serve to reduce in-school transmission by up to 30%.

Of greater significance is the comparison of different strategies to deal with infection occurrences. The exclusion of whole classrooms and bubbles has led to self-isolation of large numbers of pupils from primary schools in the recent past, with major adverse effects on education. In addition, such absences have significant and unwelcome knock-on effects on families and the economy. Our model results show that, in terms of infection numbers and transmissions, there are no tangible benefits of this policy in comparison to a policy of simply removing a pupil who has been infected from school. Thus, excluding whole classes or bubbles is hard to justify from a public health perspective and in the context if its highly disruptive effects on education and society.

Our modelling also shows that a regular rapid lateral flow testing regime has significant benefits in reducing transmission and is more effective than bubble or class exclusion. This results in much lower absence numbers, so is manifestly superior from an educational and societal perspective. Contrasts between different testing regimes with respect to frequency of testing and whether lateral flow testing is augmented by PCR tests are marginal. Our model results indeed suggest that augmented PCR tests have little benefit since a positive lateral flow result will, in most cases, simply be confirmed by a PCR test. Thus, a testing regime which meets criteria of practicality and cost should be chosen. In the context of our modelling, the findings suggest two lateral flow tests per week is a suitable regime, with little or no need for PCR tests.

Our models have limited ability to forecast. There are very large uncertainties in future projections of incidence for the forthcoming autumn. Community incidence is the main driver of infection is school. In the simulations this remains the case even with a large proportion of the adult population vaccinated. We caution though that - to some extent - community prevalence control is built into our models (see next section for fuller discussion).

Caveats, limitations and further developments

Our model makes a number of assumptions and has some limitations. We are considering only close contacts (for clarity, define here as per [1]) as an infection transmission pathway. Unavoidably, the contact pattern data we have used, derived from teacher elicitation [1], are an approximation of complex varying patterns of human interactions. Our model also assumes each classroom is isolated

and independent of other classrooms, as far as transmission likelihood is concerned. Clearly there are interactions between pupils and adult staff from different classes and across age groups, but we judge this simplification is justified from evidence of limited cross age mixing, e.g. [6].

Our model does not consider other measures to mitigate transmission (e.g., masks, cleaning and ventilation). Some or all of these might be taken into account in development of a more sophisticated, more complicated model; this said, we doubt there would be much benefit for such additional complexities, given they would all introduce uncertainties and, it can be argued, would only introduce second order factors and effects.

One major issue is that the model does not explicitly distinguish between community-related infection transmission outside and inside school. Thus, to a large extent, the current model is predicated intrinsically on community prevalence being dominant. The model is, however, able to identify transmission clusters and larger outbreaks of in-school infections. We think this aspect of modelling could be usefully developed to allow it to detect differences between infection rates inside and outside school separately. With much of the adult population vaccinated, it could be that schools become the main reservoirs and drivers of overall community infection, rather than the other way around.

In terms of 'outbreaks' of infection, at this stage we have concentrated on projections for six or more infections in a class because this breakpoint clearly emerges from the data. In this regard, in Section 4 we mention current DfE contingency operational guidance for schools [26] where two thresholds are defined for "extra action" under specific setting conditions and in relation to 'close mixing' groups within schools. If we take the second of these, i.e., "10% of children, pupils, students or staff who are likely to have mixed closely test positive for COVID-19 within a 10-day period", for our hypothetical class of thirty this equates to 3 or more positives. *En passant*, in Figure 15c, we present projected number of classes - in a collection of ten schools with four classes each - that could match this concentration of infections, under two mitigation scenarios. There is, therefore, a need to focus our modelling specifically on these stated DfE criteria, and to develop such projections – and, critically, their associated uncertainties - in more detail than has been possible thus far.

Another useful potential development relates to forecasting. Forecast reliability in such models is strongly influenced by the extent of associated uncertainties. However, our modelling might usefully be combined with near real-time data assimilation of observed incidence to generate short-term forecasts that could be reasonably dependable, out to a few weeks ahead.

7. Conclusions

Substantive implications for the relative efficacies of different mitigation measures in schools are indicated by our agent-based contact transmission modelling results, reported above. For instance, our modelling adduces evidence that, with the emergence of the Delta variant and other changing factors, there will be a distinct contrast between the efficacies of bubbles and testing when it to comes to projected numbers of future Covid-19 infections in school and consequent absences. We contend this is surely a key finding for policy considerations. In addition, our results indicate potential benefits from refocusing on social distancing in schools, as well as enhanced ventilation and good hygiene.

These, and related arguments from our modelling, suggest that - in forecast mode - our stochastic contact network framework should be of value to Directors of Public Health and Education Authorities charged with responsibility for prompt assessment and managing of infection transmission dynamics in school classes and in multiple schools in a given area. In terms of current or future DfE action criteria [e.g. 26, 32], our model could be tailored straightforwardly to their own schools' demographics and to the relevant local community incidence level.

The numerical modelling code has been tuned for rapid analytical computation and, therefore, could be used on a frequent, real-time basis under changing Covid-19 infection conditions; with suitable resources, daily data-adaptive updating of projections should be feasible for a group of schools in a region.

Acknowledgments

We thank Dr Ellen Brooks-Pollock (Bristol Veterinary School, University of Bristol), Prof. Sarah Lewis (Bristol Medical School, University of Bristol) and Prof. Caroline Relton (Bristol Medical School, University of Bristol) for their generous advice and assistance in developing this study.

MJW and WPA acknowledge funding from the "COVID-19 Mapping and Mitigation in Schools" (CoMMinS) project which is supported by the Medical Research Council (grant no. MR/V0285545/1) and hosted within the UK MRC Integrative Epidemiology Unit at the University of Bristol (MC_UU_00011/5). There was no direct funding for this study but the support of the Royal Society RAMP initiative for COVID-19 is acknowledged. MJW also acknowledges funding from NERC on fellowship NE/R003890/1.

References

[1] Sparks, R. S. J., Aspinall, W. P., Brooks-Pollock, E., *et al.* 2021 A novel approach for evaluating contact patterns and risk mitigation strategies for COVID-19 in English primary schools with application of structured expert judgement. *Royal Society Open Science*. **8**, 201566. (doi:10.1098/rsos.201566)

[2] Conlan, A. J. K., Eames, K. T. D., Gage, J. A., *et al.* 2011 Measuring social networks in British primary schools through scientific engagement. *Proceedings of the Royal Society B: Biological Sciences*. **278**, 1467-1475. (doi:10.1098/rspb.2010.1807)

[3] Ferretti, L., Ledda, A., Wymant, C., *et al.* 2020 The timing of COVID-19 transmission. *medRxiv*. 2020.2009.2004.20188516. (10.1101/2020.09.04.20188516)

[4] McAloon, C., Collins, Á., Hunt, K., *et al.* 2020 Incubation period of COVID-19: a rapid systematic review and meta-analysis of observational research. *BMJ Open*. **10**, e039652. (10.1136/bmjopen-2020-039652)

[5] Aspinall, W. P., Sparks, R. S. J., Cooke, R. M., *et al.* 2020 Quantifying threat from COVID-19 infection hazard in Primary Schools in England. *medRxiv*. 2020.2008.2007.20170035. (10.1101/2020.08.07.20170035)

[6] Davies, N. G., Klepac, P., Liu, Y., *et al.* 2020 Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nature Medicine*. **26**, 1205-1211. (10.1038/s41591-020-0962-9)

[7] COVID-19 in children and the role of school settings in transmission - first update. Stockholm: European Centre for Disease Prevention and Control 2020.

[8] He, D., Zhao, S., Lin, Q., *et al.* 2020 The relative transmissibility of asymptomatic COVID-19 infections among close contacts. *International Journal of Infectious Diseases*. **94**, 145-147. (10.1016/j.ijid.2020.04.034)

[9] Menni, C., Valdes, A. M., Freidin, M. B., *et al.* 2020 Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nature Medicine*. **26**, 1037-1040. (10.1038/s41591-020-0916-2)

[10] Molteni, E., Sudre, C. H., Canas, L. S., *et al.* 2021 Illness duration and symptom profile in a large cohort of symptomatic UK school-aged children tested for SARS-CoV-2. *medRxiv*. 2021.2005.2005.21256649. (10.1101/2021.05.05.21256649)

[11] Cooke, R. M. 1991 *Experts in uncertainty: Opinion and subjective probability in science*. New York, NY, US: Oxford University Press.

[12] Rongen, G., Hart, C. M. P., Leontaris, G., *et al.* 2020 Update (1.2) to ANDURIL and ANDURYL: Performance improvements and a graphical user interface. *SoftwareX*. **12**, 100497. (https://doi.org/10.1016/j.softx.2020.100497)

[13] van der Hofstad, R. 2016 *Random Graphs and Complex Networks*. Cambridge: Cambridge University Press.

[14] Coronavirus (COVID-19) in the UK: Vaccinations. gov.uk 2021 [cited 12 August 2021]; Available from: https://coronavirus.data.gov.uk/details/vaccinations

[15] Keeling, M., Dyson, L., Hill, E., *et al.* University of Warwick: Roadmap scenarios and sensitivity – Step 4, 6 July 2021. https://www.gov.uk/government/publications/university-of-warwick-roadmap-scenarios-and-sensitivity-step-4-6-july-2021: Scientific Advisory Group for Emergencies 2021.

[16] Sonabend, R., Imai, N., Knock, E. S., *et al.* Imperial College London: Evaluating the roadmap out of lockdown for England: modelling the delayed Step 4 of the roadmap in the context of the Delta variant, 7 July 2021 https://www.gov.uk/government/publications/imperial-college-london-evaluating-the-roadmap-out-of-lockdown-for-england-modelling-the-delayed-step-4-of-the-roadmap-in-the-context-of-the-delta-v: Scientific Advisory Group for Emergencies 2021.

[17] Barnard, R. C., Davies, N. G., Jit, M., *et al.* LSHTM: Updated roadmap assessment – prior to delayed Step 4, 7 July 2021 https://www.gov.uk/government/publications/lshtm-updated-roadmap-assessment-prior-to-delayed-step-4-7-july-2021: Scientific Advisory Group for Emergencies 2021.

[18] Evidence summary: COVID-19 - children, young people and education settings. Department for Education Education, universities and childcare during coronavirus July 2021 [cited August 2021]; Available from: https://www.gov.uk/government/publications/evidence-summary-covid-19-children-young-people-and-education-settings

[19] Thompson, H. A., Mousa, A., Dighe, A., *et al.* 2021 Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Setting-specific Transmission Rates: A Systematic Review and Meta-analysis. *Clinical Infectious Diseases*. **73**, e754-e764. (10.1093/cid/ciab100)

[20] Mulchandani, R., Jones, H. E., Taylor-Phillips, S., *et al.* 2020 Accuracy of UK Rapid Test Consortium (UK-RTC) "AbC-19 Rapid Test" for detection of previous SARS-CoV-2 infection in key workers: test accuracy study. *BMJ*. **371**, m4262. (10.1136/bmj.m4262)

[21] Deeks, J. J., Raffle, A. E. 2020 Lateral flow tests cannot rule out SARS- CoV-2 infection. *BMJ*. **371**, m4787. (10.1136/bmj.m4787)

[22] COVID-19 Infection Survey: methods and further information Office for National Statistics [cited 28 July 2020]; Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/

methodologies/covid19infectionsurveypilotmethodsandfurtherinformation#test-sensitivity-and-specificity

[23] Schools, pupils and their characteristics. Department for Education January 2021 [cited 2021; Available from: https://www.gov.uk/government/statistics/schools-pupils-and-their-characteristics-january-2021

[24] Steel, K., Davies, B. COVID-19 Infection Survey. Office for National Statistics 2020.

[25] Aspinall, W. P., Sparks, R. S. J., Woodhouse, M. J., *et al.* in press Pupils returning to Primary Schools in England during 2020: rapid estimations of punctual COVID-19 infection rates. *Royal Society Open Science*.

[26] Contingency framework: education and childcare settings, August 2021. Department for Education 2021 [cited August 2021; Available from: https://www.gov.uk/government/publications/coronavirus-covid-19-local-restrictions-in-education-and-childcare-settings

[27] Attendance in education and early years settings during the coronavirus (COVID-19) outbreak. Department for Education COVID Attendance Statistics 2020 [cited 28 July 2020]; Available from: https://explore-education-statistics.service.gov.uk/find-statistics/attendance-in-education-and-early-years-settings-during-the-coronavirus-covid-19-outbreak/2020-week-43

[28] Torjesen, I. 2021 Covid-19: Delta variant is now UK's most dominant strain and spreading through schools. *BMJ*. **373**, n1445. (10.1136/bmj.n1445)

[29] SARS-CoV-2 variants of concern and variants under investigation in England: technical briefing 20. Public Health England 2021 [cited; Available from: https://www.gov.uk/government/publications/investigation-of-novel-sars-cov-2-variant-variant-of-concern-20201201

[30] Alizon, S., Haim-Boukobza, S., Foulongne, V., *et al.* 2021 Rapid spread of the SARS-CoV-2 Delta variant in some French regions, June 2021. *Eurosurveillance*. **26**, 2100573. (doi:https://doi.org/10.2807/1560-7917.ES.2021.26.28.2100573)

[31] Li, B., Deng, A., Li, K., *et al.* 2021 Viral infection and transmission in a large, well-traced outbreak caused by the SARS-CoV-2 Delta variant. *medRxiv*. 2021.2007.2007.21260122. (10.1101/2021.07.07.21260122)

[32] Schools COVID-19 operational guidance. Department for Education 17 August 2021 [cited 25 August 2021]; Available from: https://www.gov.uk/government/publications/actions-for-schools-during-the-coronavirus-outbreak/schools-covid-19-operational-guidance

[33] Keeling, M. J., Rohani, P. 2011 *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press.

Supplementary Material

A1. Linking prevalence to incidence rate

The roadmap projections [15-17] that we use to forecast potential future infections in schools contain trajectories for the incidence rates through Term 1 of 2021, but our model also requires time series for the community prevalence. Here we adopt a simple relationship between incidence rate and prevalence, as described here.

Our relationship is based on the SIR compartmental epidemiological model [33] that relates the numbers of susceptible (S), infected (I) and recovered (R) people in the population. New infections occur at a rate depending on the number of contacts between susceptible and infected people, with a parameter β (given by the product of the contact rate and probability of transmission in a contact). Infected people recover (or are otherwise removed) with probability of recovery in time interval dt given by γdt (which is equivalent to assuming duration of infection is drawn from an exponential distribution, with mean duration $D = 1/\gamma$).

The movement of people between compartments is encoded in a system of differential equations [33], given by

$$\frac{dS}{dt} = -\beta\frac{SI}{N}, \quad \frac{dI}{dt} = \beta\frac{SI}{N} - \gamma I, \quad \frac{dR}{dt} = \gamma I$$

where $N = S + I + R$ is the total population (here assumed constant). The second of these equations is pertinent to understanding the link between incidence rate and prevalence. Indeed, this equation can be written in the form

$$\frac{dI}{dt} = \left(R_0\frac{S}{N} - 1\right)\gamma I,$$

where R_0 is the basis reproduction number, and where I is the community prevalence and dI/dt is the incidence rate of new infections in the population. For SARS-CoV-2, we simplify this expression further by noting that the data suggests that the infected population at any time is a relatively small proportion of the total population in the UK, and that recovery does not appear to prevent reinfection. Therefore we can make the approximation that $S/N \approx 1$, so that

$$\frac{dI}{dt} \approx (R_0 - 1)\gamma I.$$

If the incidence rate dI/dt is known, and there are estimates of R_0 and $\gamma = 1/D$ then we can directly estimate the prevalence as

$$I = \frac{dI/dt}{(R_0 - 1)\gamma}.$$

This relationship is a simple rescaling of incidence rate to obtain prevalence and suggests that high incidence rates occur for large values of the reproduction number R_0 for a given prevalence (as expected), and that high prevalence occurs at a given incidence rate if the duration of infection is increased (as expected).

In using the expression derived above, we assume that R_0 and γ are constants. This means that contact rates are unchanged, so in analysis of past data we must take periods where social contact restrictions are unchanged. Additionally, the probability of transmission in a contact must be constant, so it is unlikely to apply in periods where there are multiple variants of SARS-CoV-2 competing within a population.

We test this relationship using data from the Office for National Statistics Coronavirus Infection Survey [24]. The methods used for the estimate in this dataset have changed over the course of the epidemic, and therefore the interpretation of the parameter γ may change. In 2020, the incidence and prevalence estimates were mainly derived from random PCR testing surveys, and therefore the duration of

infection $D = 1/\gamma$ is likely to refer to the duration for which PCR testing detects virus in a sample. This may be substantially longer than the duration of symptoms or the duration of infectivity. Figure 16 shows the relationship between prevalence and incidence over the period 1 September to 22 November 2020. In the early part of this period (up to early-October 2020), incidence and prevalence increased and a clear linear relationship is observed. In this time, some social mixing restrictions were introduced, but do not substantially alter the trend. However, from early October 2020, there was a marked increase in the prevalence with only small changes in the incidence rate which continued until a second national lockdown was announced on 31 October 2020 and enacted on 5 November 2020. While this increased prevalence could be explained by reductions in R_0 while the incidence rate remained steady, this seems unlikely, and may be better explained by a breakdown of the simple modelling assumptions during a period of change in social behaviour, e.g., a lag between prevalence and incidence that is not included in the simple SIR model. Additionally, the Alpha variant (B.1.1.7) was known to be circulating in the UK in this time.

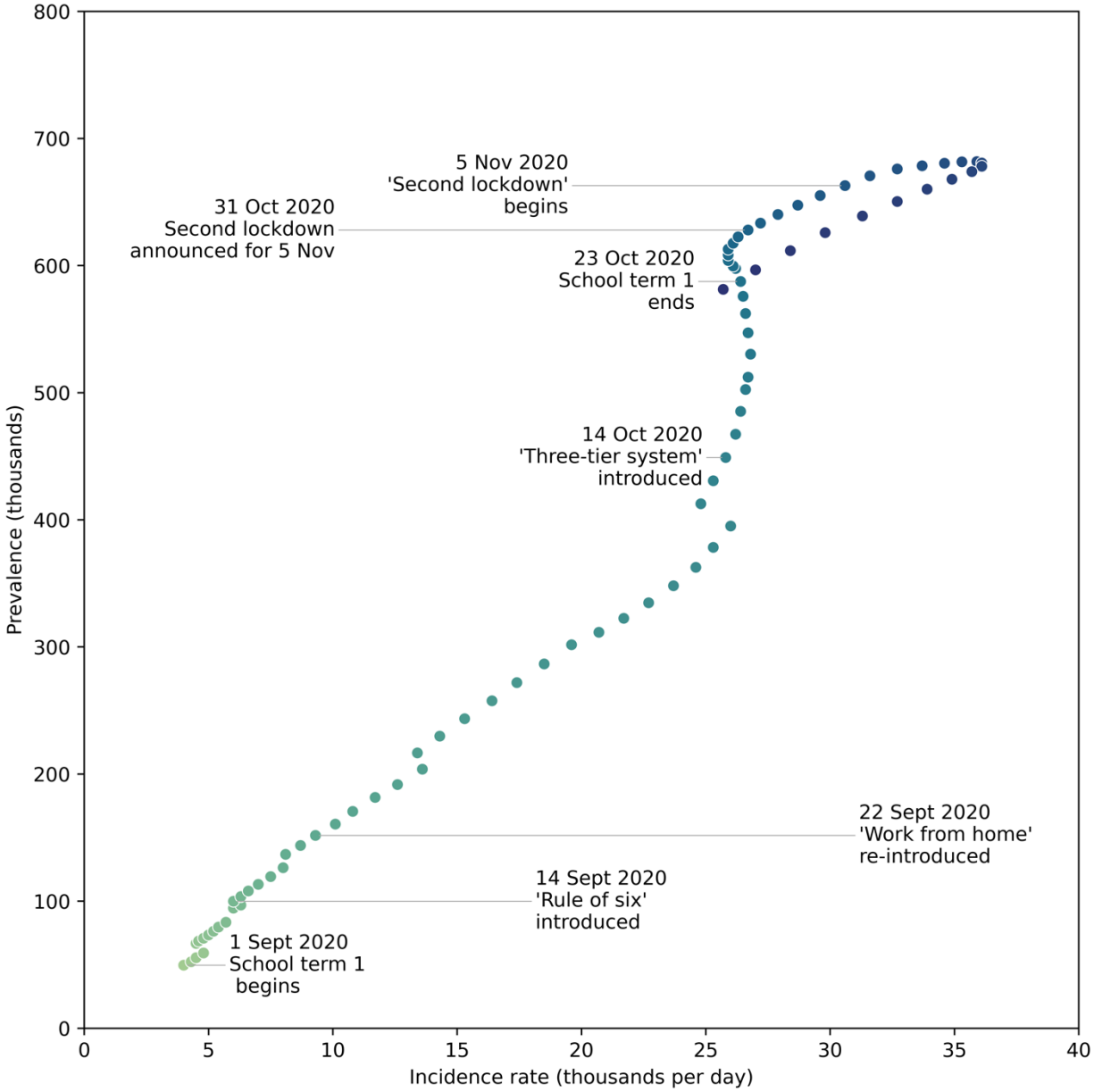

Figure 16. Estimates of the UK SARS-CoV-2 prevalence as a function of incidence rate for 1 September 2020 to 22 November 2020 (date from [24]). The colour of points indicates progression through time. Key dates for changes in social contact rates are indicated. Note, the Alpha variant (B.1.1.7) was first detected in a PCR sample taken in mid-September 2020 and may have been spreading rapidly in this time.

Despite these limitations, the rescaling relationship is a pragmatic and practical tool for estimating prevalence from incidence during periods where social contacts are not changing substantially. We therefore adopt this for estimating prevalence from projections of incidence rate for our forecasts in

Term 1 of 2021. However, this requires estimation of the R_0 and γ parameters. While a Bayesian approach would be valuable to quantify uncertainties, for our purposes we are content using a simple manual fitting procedure, using estimates of R_0 for the dominant Delta variant (given in the range 5—7 in [16]) and noting that increases in this value can be compensated by decreases in γ. This approach produces Figure 11 from ONS data for June—August 2021 and the incidence rate forecasts of [15].